4-8-2020

On Brownian Motion Governed by Telegraph Process

Chaoran Hu
University of Connecticut - Storrs, chaoran.hu@uconn.edu

On Brownian Motion Governed by Telegraph Process

Chaoran Hu, Ph.D.
University of Connecticut, 2020

ABSTRACT

Statistical modeling of animal movement is of critical importance in addressing fundamental questions about space use, movement, resource selection, and behaviour in animal ecology. The explosion of telemetry data on animal movement from the recent advancements in tracking and observation technologies presents a storm of opportunities and challenges.

A moving-resting-handling (MRH) process is introduced to allow predators to have two different motionless states, resting and handling. In essence, it is a Brownian motion whose infinitesimal variance changes according to a three-state continuous-time Markov Chain. The Markov Chain can be viewed as a telegraph process with one on state and two off states. We derive the distribution of the occupation times of this Markov Chain and develop a maximum likelihood estimation procedure when the stochastic process at hand is observed at discrete, possibly irregularly spaced time points. The likelihood function is evaluated with forward algorithm in the general framework of hidden Markov models.

Chaoran Hu, University of Connecticut, 2020

The second objective of this thesis is to introduce measurement errors to a recently proposed moving-resting (MR) process. MR process is a Brownian motion governed by a telegraph process, which allows periods of inactivity in one state of the telegraph process. It is promising in modeling the movements of predators with long inactive periods such as mountain lions, but the lack of accommodation of measurement errors seriously prohibits its applications in practice. Here we incorporate measurement errors in the MR model and derive basic properties of the model.

Finally, convolutions of independent gamma variables are encountered in many applications, including animal movement modeling. Accurate and fast evaluations of their density and distribution functions are critical for such applications. We review several numerical evaluations of convolution of independent gamma variables and compare them with respect to their accuracy and speed. We also derive a new computationally efficient formula for the probability mass function of the number of renewals by a given time.

Two R packages `smam`, `coga` provide efficient C++ based implementations of the discussed methods and are available in CRAN.

On Brownian Motion Governed by Telegraph Process

Chaoran Hu

B.S., Huazhong University of Science and Technology, Wuhan, China, 2015

M.S., University of Connecticut, CT, USA, 2017

A Dissertation
Submitted in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy
at the
University of Connecticut

2020

APPROVAL PAGE

Doctor of Philosophy Dissertation

On Brownian Motion Governed by Telegraph Process

Presented by

Chaoran Hu, B.S. Mathematics, M.S. Statistics

Major Co-Advisor _____

Dr. Jun Yan

Major Co-Advisor _____

Dr. Vladimir Pozdnyakov

Associate Advisor _____

Dr. Richard Vitale

Associate Advisor _____

Dr. Zhiyi Chi

Associate Advisor _____

Dr. Haim Bar

University of Connecticut

2020

Dedication

This dissertation is dedicated to

my parents, Mi Kong and Rendao Hu; to my grandmother, Aihua Jia.

Acknowledgements

I would like to express my deep appreciation and gratitude to my advisors, Dr. Jun Yan and Dr. Vladimir I. Pozdnyakov, for their patient guidance and detailed mentorship. I would like to pay my special regards to Dr. James J. Grady, who is my advisor in Biostatistics Center, UConn Health. I also would like to thank my committee members, Dr. Richard A. Vitale, Dr. Zhiyi Chi and Dr. Haim Y. Bar. In addition, I would like to extend my sincere thanks to Dr. Thomas H. Meyer and Dr. L. Mark Elbroch, for their hard work in data collection and expertise in ecology. I wish to acknowledge the support and great love of my mother, Mi Kong; my father, Rendao Hu; and my grandmother, Aihua Jia. I am also thankful to my friends, Dr. Rong Wu, Xinpeng Deng and Jinjian Mu.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Random walks on a plane, whether simple, biased, or correlated, have a long history of being employed by ecologists to model the movement of animals, micro-organisms, and cells on a small time scale. By the functional Central Limit Theorem, from an appropriate distance any random walk (under some mild regularity conditions) looks like a Brownian Motion (BM). So, it is not surprising that recently diffusions are often used to model animal movement on a large time scale (e.g., Preisler et al., 2004; Tilles and Petrovskii, 2016). An excellent review on applications of random walks and diffusions in this area of research can be found in Codling et al. (2008).

Horne et al. (2007) introduced the Brownian bridge movement model (BBMM) that, in essence, assumes that animal movement is perpetual and described by a BM. Pauses in animal movement (on a small time scale) were first introduced in Othmer et al. (1988) where the dispersal of cells or organisms is modeled by a process that comprises a sequence of alternating pauses and jumps. The moving-resting (MR) process introduced in Yan et al. (2014) and further investigated in Pozdnyakov et al. (2019) allows an animal to have two states, moving and resting. In the moving state, the motion is characterized

by a BM; in the resting state, there is no movement. The duration in either moving or resting states is assumed to be exponentially distributed.

Properties and fitting of the MR model are based on results for telegraph processes (the alternating renewal process or the on-off process) that were obtained in Perry et al. (1999), Di Crescenzo (2001), Stadje and Zacks (2004), and Zacks (2004). The distribution of total time spent in a state plays a critical role in applications driven by a telegraph process (Zacks, 2012). In particular, a BM governed by a telegraph process is an active area of research such as being recently employed in continuous-time option pricing theory (e.g., Di Crescenzo and Pellerey, 2002; Kolesnik and Ratanov, 2013; Di Crescenzo et al., 2014; Di Crescenzo and Zacks, 2015).

In this dissertation, we extend the moving-resting model in two directions. The first extension is to introduce of an additional motionless state. This allows us to produce a more realistic from a biological point of view animal movement model. The second objective is to develop an efficient composite likelihood estimation procedure when the observations have added measurement errors. Incorporation of measurement errors is an important feature of existing animal movement models.

First, let us provide motivation for the first extension. It is reasonable to assume that there are very different explanations for why a predator is not moving. For example, an animal might spend time resting (as in Yan et al., 2014), consuming a prey item, or denning. Resting can be assumed to not last even a single day. However, some predators that can kill a (relatively) large prey item evolved highly elastic guts, and they consume

the kill by repeatedly gorging and digesting over a prolonged period called *handling*. For example, mountain lions (*Puma concolor*) might remain at a kill for days. Both resting and handling are periodic in the time scales of this model but denning is not, and it is inapplicable to male mountain lions in any case. Therefore, this model concerns only two non-moving activities, resting and handling, and it is clear that their durations must be different.

In the new model we have one moving state and two motionless states. From a motionless state one always switches to the moving state. Nonetheless, when moving ends, the motionless state type is chosen randomly. For tractability, all the durations (or holding times) are exponentially distributed. We will call this continuous-time process a *moving-resting-handling process*, or *MRH process*. An extension of the telegraph process to an alternating process with three states is studied in Bshouty et al. (2012). The difference is that in Bshouty et al. (2012) three states alternate deterministically within a renewal cycle. In our case we have only two states within a renewal cycle but one of the motionless states is chosen at random.

In practice, an MRH process is typically observed at discrete, possibly irregularly spaced time points. Estimation of MRH process parameters is challenging because the states are unobserved, and the observed sequence is not Markov. Our estimation procedure uses techniques developed for the hidden Markov model (HMM). More specifically, the dynamic programming, or the forward algorithm, for HMMs is employed to

compute the likelihood (e.g., Cappé et al., 2005). As will be seen, the key to this problem is the distribution of the time that the MRH process spends in the moving state. Our methodology differs from the standard approach to occupation time distribution in continuous-time Markov chain (Sericola, 2000). The method is general so that it remains valid when the holding times are not exponentially distributed, in which case, the state process is semi-Markov.

The second extension deals with the issue of measurement errors. One major limitation in applying the MR model to animal movement data is that it does not accommodate the measurement error of telemetric devices. Adding random noise to a Brownian motion model is not crucial as long as the measurement error is small in comparison to the total standard deviation of the increments of the Brownian motion between two consecutive time points (Pozdnyakov et al., 2014). In such cases, discarding the measurement error would not produce significant bias. The impact of the measurement errors on inferences about MR processes, however, is much greater. For a given sequence of hidden states, the likelihood is a product of both densities *and* probabilities. If two observed locations are exactly the same, that is, the trajectory is flat over time, then the animal is known to be motionless between the two time points. The likelihood contribution is the probability of staying in the motionless state instead of the density of the increment at zero. Adding even a tiny bit of noise would remove those flat pieces of the trajectory and, hence, cause drastic bias in the likelihood estimator of the parameters. One possible approach is to round the observed locations, which introduces flat pieces. The number of

such pieces, however, depends greatly on the rounding level, which there are no obvious rules to choose. A detailed illustration of the issue is provided in the Chapter 4.

Dealing with added noise in the MR process is challenging because it invalidates the Markov property of the joint location-state process. The transition density from one time to the next time point can in principle be obtained from convoluting the results of the MR process (Yan et al., 2014) with normally distributed measurement errors, although computationally very intensive. Lack of Markov property of the joint location-state process means the true likelihood cannot be easily formed by multiplying these transition densities. Because the measurement errors are continuous, the dynamic programing tools of HMM based on a finite number of hidden states (Cappé et al., 2005) are not applicable. The generic simulation based inferences such as iterated filtering (Ionides et al., 2011, 2015) or particle Markov chain Monte Carlo (Andrieu et al., 2010), available in R package `pomp` (King et al., 2016) are not feasible from our investigation due to the complexity of the MR process with measurement error.

Our contribution is a toolbox for applying the MR process with measurement error to animal movement modeling. First, we show that discarding the measurement error, even tiny bit ones, cause severe bias in estimation, and that rounding does not provide any satisfactory solution. To make inferences for MR process with measurement error, we establish that, after thinning every other observation, the remaining observations are location-state Markov. This facilitates a composite likelihood which contains two true likelihood components, one based on odd-numbered observations and the other based on

even-numbered observations. The true likelihood of each component is computed with dynamic programming. The variance of the maximum composite likelihood estimator can be estimated by a sandwich variance estimator or through parametric bootstrap. The validity of the approach is confirmed in a simulation study. We then apply the approach to model the movement data of mountain lion in Wyoming, whose trajectory is known to have long inactive periods.

Finally, accurate and fast evaluation of the density and distribution function of convolution of independent gamma variables is important in many applications. Examples are total claims of insurance policies (Kaas et al., 2008, chapter 2), total risks of a capital portfolio (Furman and Landsman, 2005), storage capacity measurement of a dam (Mathai, 1982), lifetimes of redundant standby systems with independent components in reliability analysis (Bon and Pãltãnea, 1999; Jasiulewicz and Kordecki, 2003; Kadri et al., 2015), and total inter-arrival times of network traffic (Nadarajah and Kotz, 2007) or, more generally, a point process (Sim, 1992). Convolution of independent gamma variables also plays an important role in both moving-resting-handling process and moving-resting process with measurement error model. We discussed this topic in Appendix A.

An implementation of all methods in this dissertation and in our earlier works (Pozdnyakov et al., 2014; Yan et al., 2014; Pozdnyakov et al., 2019), is publicly available in an R package `smam` (Hu et al., 2020b). All methods related to convolution of independent gamma variables are implemented as an R package `coga` (Hu et al., 2019).

This dissertation is arranged as follows. In Chapter 2, we introduce the telegraph process with multiple off-states. Then, we derive the marginal distribution of occupation time of one off-state and the joint distribution of occupation times of two off-states. In Chapter 3, we study moving-resting-handling process (the Brownian motion governed by telegraph process with multiple off-states). The forward/backward algorithm and Viterbi algorithm are constructed to calculate likelihood and to address the state prediction problem. In Chapter 4, we introduce measurement errors to the MR model. Two computationally efficient composite likelihood estimation procedures are developed. The question of standard errors of estimates is also addressed. The numerical evaluation of convolution of independent gamma variables are discussed in details in Appendix A.

Chapter 2

On-Off Processes with Multiple Off-States

2.1 Algorithmic Construction of On-Off Process with Two Off-States

Suppose that we are given the following collection of independent sequences of non-negative random variables:

1. $\{M_k\}_{k\geq 1}$ are independent identically distributed (iid) random variables with cumulative distribution function (cdf) F_M and probability density function (pdf) f_M (these random variables will be used as the holding times when the server is up, state 0);

2. $\{R_k\}_{k\geq 1}$ are iid random variables with cdf F_R and pdf f_R (holding times when the server is down for short repairs, state 1);

3. $\{H_k\}_{k\geq 1}$ are iid random variables with cdf F_H and pdf f_H (holding times when the

server is down for long repairs, state 2);

4. $\{\xi_k\}_{k\geq 1}$ are iid random variables with $\Pr(\xi_k = 1) = p_1 > 0$ and $\Pr(\xi_k = 0) = p_2 = 1 - p_1$.

Now, we present our construction of the on-off process with two off-states, $S(t)$.

1. Initialize with $S(0) = 0$ and $T_0 = 0$.

2. For cycles $i = 1, 2, \ldots$:

   (a) Let $T_{2i-1} = M_i + T_{2i-2}$, and $S(t) = 0$ for all $t \in [T_{2i-2}, T_{2i-1})$.

   (b) If $\xi_i = 1$ then $T_{2i} = T_{2i-1} + R_i$, and $S(t) = 1$ for all $t \in [T_{2i-1}, T_{2i})$; otherwise, $T_{2i} = T_{2i-1} + H_i$, and $S(t) = 2$ for all $t \in [T_{2i-1}, T_{2i})$.

Then the occupation times in state 0, state 1, and state 2 are, respectively,

$$M(t) = \int_0^t 1_{\{S(s)=0\}} \mathrm{d}s,$$
$$R(t) = \int_0^t 1_{\{S(s)=1\}} \mathrm{d}s,$$
$$H(t) = \int_0^t 1_{\{S(s)=2\}} \mathrm{d}s.$$

The corresponding defective marginal densities of $R(t)$ and $H(t)$ are denoted as

$$p_{Rj}(s,t) = \Pr(R(t) \in \mathrm{d}s, S(t) = j)/\mathrm{d}s,$$

$$p_{Hj}(s,t) = \Pr(H(t) \in \mathrm{d}s, S(t) = j)/\mathrm{d}s,$$

where $t \geq 0$, $0 < s < t$, $j = 0, 1, 2$. The defective joint two-dimensional density of $R(t)$ and $H(t)$ for $u, v > 0$, $u + v < t$ is denoted as

$$p_{RHj}(u, v, t) = \frac{1}{\mathrm{d}u\mathrm{d}v} \Pr(R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, S(t) = j),$$

where $j = 0, 1, 2$. The densities are defective in the following sense: since both occupation times $R(t)$ and $H(t)$ have an atom at 0, all three sums

$$\sum_{j=0}^{2} \int_0^t p_{Rj}(s, t)\mathrm{d}s, \quad \sum_{j=0}^{2} \int_0^t p_{Hj}(s, t)\mathrm{d}s,$$

and

$$\sum_{j=0}^{2} \iint_{u,v>0,\, u+v<t} p_{RHj}(u, v, t)\mathrm{d}u\mathrm{d}v$$

are less than 1.

2.2 Marginal Distribution of Occupation Time of an Off-State

To derive the distribution of occupation time we will need some auxiliary random variables. Let $N(t)$ be the number of cycles (or returns to the on-state) by time t. Formally, for $n \geq 0$

$$N(t) = n \qquad \text{iff} \qquad T_{2n} \leq t < T_{2n+2}. \tag{2.1}$$

Finally, let $U_k = \xi_k R_k + (1 - \xi_k)H_k$. These random variables are associated with the off-state holding time of the regular on-off process when two off-states are combined.

Our formulas include convolutions of different distributions. We will use the following notation. If we are given a cdf $G(\cdot)$ and its pdf $g(\cdot)$, then $G^{(n)}(\cdot)$ denotes n-fold convolution of $G(\cdot)$, and $g^{(n)}(\cdot)$ denotes the n-fold convolution of $g(\cdot)$. If we are given two cdfs $G(\cdot)$ and $H(\cdot)$ with pdfs $g(\cdot)$ and $h(\cdot)$, then $G * H(\cdot)$ denotes the convolution of cdfs $G(\cdot)$ and $H(\cdot)$, and $g * h(\cdot)$ denotes the convolution of pdfs $g(\cdot)$ and $h(\cdot)$.

We also use the following conventions. Any summation over the empty set is 0. Zero-fold convolution $G^{(0)}(\cdot)$ is the cdf of a random variable that is equal to 0 with probability 1. Finally, $g^{(k)} * h^{(0)}(\cdot) = g^{(k)}(\cdot)$ for any $k \geq 1$.

Theorem 1. *Let $t \geq 0$ and $0 < s < t$. Then*

$$\Pr(R(t) = 0, S(t) = 0) = \sum_{n=0}^{\infty} \left[F_M^{(n)} * F_H^{(n)}(t) - F_M^{(n+1)} * F_H^{(n)}(t) \right] p_2^n,$$

$$\Pr(R(t) = 0, S(t) = 1) = 0,$$

$$\Pr(R(t) = 0, S(t) = 2) = \sum_{n=0}^{\infty} \left[F_M^{(n+1)} * F_H^{(n)}(t) - F_M^{(n+1)} * F_H^{(n+1)}(t) \right] p_2^{n+1},$$

and

$$p_{R0}(s,t) = \sum_{n=1}^{\infty} \sum_{k=1}^{n} \binom{n}{k} p_1^k p_2^{n-k} f_R^{(k)}(s) \times \left[F_M^{(n)} * F_H^{(n-k)}(t-s) - F_M^{(n+1)} * F_H^{(n-k)}(t-s) \right],$$

$$p_{R1}(s,t) = \sum_{n=0}^{\infty}\sum_{k=0}^{n}\binom{n}{k}p_1^{k+1}p_2^{n-k}f_M^{(n+1)} * f_H^{(n-k)}(t-s)\left[F_R^{(k)}(s) - F_R^{(k+1)}(s)\right],$$

$$p_{R2}(s,t) = \sum_{n=1}^{\infty}\sum_{k=1}^{n}\binom{n}{k}p_1^{k}p_2^{n-k+1}f_R^{(k)}(s)$$
$$\times\left[F_M^{(n+1)} * F_H^{(n-k)}(t-s) - F_M^{(n+1)} * F_H^{(n-k+1)}(t-s)\right].$$

Proof. First, note that the distribution of $R(t)$ has an atom. Indeed, if $M_1 > t$ or all the failures that occur before t are of the second type, then $R(t) = 0$. More specifically, by conditioning on the number of returns to the on-state, $N(t)$, we obtain that

$$\begin{aligned}
&\Pr(R(t)=0, S(t)=0)\\
&= \sum_{n=0}^{\infty}\Pr\left(R(t)=0, S(t)=0, N(t)=n\right)\\
&= \Pr(M_1>t) + \sum_{n=1}^{\infty}\Pr\left(\sum_{j=1}^{n}(M_j+H_j)\le t, \sum_{j=1}^{n}(M_j+H_j)+M_{n+1}>t, \sum_{j=1}^{n}\xi_j = 0\right)\\
&= \Pr(M_1>t) + \sum_{n=1}^{\infty}\Pr\left(\sum_{j=1}^{n}(M_j+H_j)\le t, \sum_{j=1}^{n}(M_j+H_j)+M_{n+1}>t\right)p_2^n\\
&= \Pr(M_1>t) + \sum_{n=1}^{\infty}\left[\Pr\left(\sum_{j=1}^{n}M_j+\sum_{j=1}^{n}H_j\le t\right) - \Pr\left(\sum_{j=1}^{n+1}M_j+\sum_{j=1}^{n}H_j\le t\right)\right]p_2^n\\
&= (1-F_M(t)) + \sum_{n=1}^{\infty}\left[F_M^{(n)} * F_H^{(n)}(t) - F_M^{(n+1)} * F_H^{(n)}(t)\right]p_2^n\\
&= \sum_{n=0}^{\infty}\left[F_M^{(n)} * F_H^{(n)}(t) - F_M^{(n+1)} * F_H^{(n)}(t)\right]p_2^n.
\end{aligned}$$

Next, again by conditioning on $N(t)$, we get that for $0 < s < t$

$$\Pr(R(t) \in \mathrm{d}s, S(t) = 0) = \sum_{n=1}^{\infty} \Pr\left(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = n\right).$$

Note that the summation starts from 1, because $S(t) = 0$ and $N(t) = 0$ implies that $M_1 > t$, and, therefore, $R(t) = 0$.

The next step is to fix the number of switches to failures of type 1. Since the total numbers of switches is n and there is at least one failure of type 1, we have

$$\begin{aligned}&\Pr(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = n)\\ &\qquad = \sum_{k=1}^{n} \Pr\left(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = n, \sum_{j=1}^{n} \xi_j = k\right).\end{aligned}$$

Because

$$S(t) = 0, N(t) = n \qquad \text{iff} \qquad \sum_{j=1}^{n}(M_j + U_j) \le t, \sum_{j=1}^{n}(M_j + U_j) + M_{n+1} > t,$$

and it really does not matter during which cycles the switches to failures of type 1 occur, we find that

$$\begin{aligned}&\Pr\left(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = n, \sum_{j=1}^{n} \xi_j = k\right)\\ &= \Pr\left(R(t) \in \mathrm{d}s, \sum_{j=1}^{n}(M_j + U_j) \le t, \sum_{j=1}^{n}(M_j + U_j) + M_{n+1} > t, \sum_{j=1}^{n} \xi_j = k\right)\end{aligned}$$

$$= \binom{n}{k} \Pr\left(R(t) \in \mathrm{d}s, \sum_{j=1}^{n} (M_j + U_j) \leq t, \sum_{j=1}^{n} (M_j + U_j) + M_{n+1} > t, \right.$$
$$\left. \sum_{j=1}^{k} \xi_j = k, \sum_{j=k+1}^{n} \xi_j = 0 \right).$$

Next, observe that in this case $R(t) = \sum_{j=1}^{k} R_j$. Using independence between $\{\xi_k\}_{k\geq 1}$ and the holding time sequences we have

$$\Pr\left(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = 0, \sum_{j=1}^{n} \xi_j = k \right)$$
$$= \binom{n}{k} p_1^k p_2^{n-k} \Pr\left(\sum_{j=1}^{k} R_j \in \mathrm{d}s, \sum_{j=1}^{n} M_j + \sum_{j=1}^{k} R_j + \sum_{j=k+1}^{n} H_j \leq t, \right.$$
$$\left. \sum_{j=1}^{n+1} M_j + \sum_{j=1}^{k} R_j + \sum_{j=k+1}^{n} H_j > t \right)$$
$$= \binom{n}{k} p_1^k p_2^{n-k} \Pr\left(\sum_{j=1}^{k} R_j \in \mathrm{d}s, \sum_{j=1}^{n} M_j + \sum_{j=k+1}^{n} H_j \leq t - s, \right.$$
$$\left. \sum_{j=1}^{n+1} M_j + \sum_{j=k+1}^{n} H_j > t - s \right).$$

Finally, independence of holding time sequences gives us that

$$\Pr\left(\sum_{j=1}^{k} R_j \in \mathrm{d}s, \sum_{j=1}^{n} M_j + \sum_{j=k+1}^{n} H_j \leq t - s, \sum_{j=1}^{n+1} M_j + \sum_{j=k+1}^{n} H_j > t - s \right)$$
$$= \Pr\left(\sum_{j=1}^{k} R_j \in \mathrm{d}s \right) \left[\Pr\left(\sum_{j=1}^{n} M_j + \sum_{j=k+1}^{n} H_j \leq t - s \right) \right.$$
$$\left. - \Pr\left(\sum_{j=1}^{n+1} M_j + \sum_{j=k+1}^{n} H_j \leq t - s \right) \right]$$

$$= f_R^{(k)}(s) \left[F_M^{(n)} * F_H^{(n-k)}(t-s) - F_M^{(n+1)} * F_H^{(n-k)}(t-s) \right] \mathrm{d}s.$$

Now let us consider the case when $S(t) = 1$ (at time t the server is down for a short repair). One difference is that there is no atom in this case. As before, for $0 < s < t$ we have

$$\begin{aligned}
&\Pr(R(t) \in \mathrm{d}s, S(t) = 1) \\
&= \sum_{n=0}^{\infty} \Pr\left(R(t) \in \mathrm{d}s, S(t) = 1, N(t) = n\right) \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \Pr\left(R(t) \in \mathrm{d}s, S(t) = 1, N(t) = n, \sum_{j=1}^{n} \xi_j = k, \xi_{n+1} = 1 \right) \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} \Pr\Bigg(R(t) \in \mathrm{d}s, S(t) = 1, N(t) = n, \\
&\qquad\qquad \sum_{j=1}^{k} \xi_j = k, \sum_{j=k+1}^{n} \xi_j = 0, \xi_{n+1} = 1 \Bigg).
\end{aligned}$$

Note that event

$$\left\{ R(t) \in \mathrm{d}s, S(t) = 1, N(t) = n, \sum_{j=1}^{k} \xi_j = k, \sum_{j=k+1}^{n} \xi_j = 0, \xi_{n+1} = 1 \right\}$$

can be rewritten as

$$\left\{ R(t) \in \mathrm{d}s, \sum_{j=1}^{n+1} M_j + \sum_{j=1}^{k} R_j + \sum_{j=k+1}^{n} H_j \leq t, \right.$$

$$\sum_{j=1}^{n+1} M_j + \sum_{j=1}^{k} R_j + \sum_{j=k+1}^{n} H_j + R_{n+1} > t,$$
$$\sum_{j=1}^{k} \xi_j = k,\ \sum_{j=k+1}^{n} \xi_j = 0,\ \xi_{n+1} = 1\Big\}.$$

Since $R(t) = t - \sum_{j=1}^{n+1} M_j - \sum_{j=k+1}^{n} H_j$, we finally obtain

$$\begin{aligned}
&\Pr(R(t) \in \mathrm{d}s, S(t) = 1) \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} p_1^{k+1} p_2^{n-k} \\
&\qquad \times \Pr\left(\sum_{j=1}^{n+1} M_j + \sum_{j=k+1}^{n} H_j \in t - \mathrm{d}s, \sum_{j=1}^{k} R_j \le s, \sum_{j=1}^{k} R_j + R_{n+1} > s \right) \\
&= \sum_{n=0}^{\infty} \sum_{k=0}^{n} \binom{n}{k} p_1^{k+1} p_2^{n-k} f_M^{(n+1)} * f_H^{(n-k)}(t-s) \left[F_R^{(k)}(s) - F_R^{(k+1)}(s) \right] \mathrm{d}s.
\end{aligned}$$

The other formulas can be derived in a similar way. □

Theorem 1 gives us the distribution of occupation time in state 1. Since both off-states enter our story in a completely symmetric way, the formulas for the occupation time in state 2 can be obtained by interchanging state 1 and 2 in Theorem 1.

2.3 Joint Distribution of Off-State Occupation Times

As mentioned in the introduction, if we are interested in the distribution of the total cost, then we need the joint distribution of off-state occupation times. More specifically,

assume that the total cost $C(t)$ is a linear function of occupation times, that is,

$$C(t) = C_0 M(t) + C_1 R(t) + C_2 H(t).$$

Then the distribution of $C(t)$ is fully determined by the joint distribution of $R(t)$ and $H(t)$. Note that we do not need the joint distribution of all three occupation times, because $M(t) = t - R(t) - H(t)$.

Since both $R(t)$ and $H(t)$ have atoms at 0, for every value of $S(t)$ we have four cases: (1) both occupation times are 0; (2) $R(t) = 0$, $H(t) > 0$; (3) $R(t) > 0$, $H(t) = 0$; and (4) both $R(t)$ and $H(t)$ are strictly greater than 0. In total, we have 12 formulas. Some of them are trivial. For instance, event $\{R(t) = 0, S(t) = 1\}$ has probability 0, therefore, the corresponding defective one-dimensional density of $H(t)$ is also 0. Moreover, since there is a certain symmetry between $R(t)$ and $H(t)$ some formulas can be found by interchanging state 1 and 2. That is why in the following theorem below we have only 5 formulas.

Theorem 2. *Let $u, v > 0$ and $u + v < t$. Then*

$$\Pr(R(t) = 0, H(t) = 0, S(t) = 0) = 1 - F_M(t),$$

$$\Pr(R(t) = 0, H(t) \in \mathrm{d}v, S(t) = 0)/\mathrm{d}v$$

$$= \sum_{n=1}^{\infty} f_H^n(v) \left[F_M^{(n)}(t - v) - F_M^{(n+1)}(t - v) \right] p_2^n,$$

$$\Pr(R(t) \in \mathrm{d}u, H(t) = 0, S(t) = 1)/\mathrm{d}u$$
$$= \sum_{n=0}^{\infty} p_1^{n+1} f_M^{(n+1)}(t-u) \left[F_R^{(n)}(u) - F_R^{(n+1)}(u) \right],$$

and

$$p_{RH0}(u, v, t)$$
$$= \sum_{n=1}^{\infty} \sum_{k=1}^{n-1} \binom{n}{k} p_1^k p_2^{n-k} f_R^{(k)}(u) f_H^{(n-k)}(v) \left[F_M^{(n)}(t-u-v) - F_M^{(n+1)}(t-u-v) \right],$$

$$p_{RH1}(u, v, t)$$
$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} \binom{n}{k} p_1^{k+1} p_2^{n-k} f_M^{(n+1)}(t-u-v) f_H^{(n-k)}(v) \left[F_R^{(k)}(u) - F_R^{(k+1)}(u) \right].$$

Proof. We will only derive the formula for $p_{RH1}(u, v, t)$. The remaining formulas can be obtained by similar modifications of the proof of Theorem 1.

As before, we start with partitioning with respect to events $\{N(t) = n\}$. More specifically, for $u, v > 0$ and $0 < u + v < t$ we have

$$p_{RH1}(u, v, t)\mathrm{d}u\mathrm{d}v = \sum_{n=0}^{\infty} \Pr\left(R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, S(t) = 1, N(t) = n\right)$$
$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} \Pr\left(R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, S(t) = 1, N(t) = n, \sum_{j=1}^{n} \xi_j = k, \xi_{n+1} = 1 \right)$$
$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} \binom{n}{k} \Pr\Bigg(R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, S(t) = 1, N(t) = n,$$
$$\sum_{j=1}^{k} \xi_j = k, \sum_{j=k+1}^{n} \xi_j = 0, \xi_{n+1} = 1 \Bigg).$$

Note that now the upper limit of the inner summation is $n-1$, because $\sum_{j=1}^{n}\xi_j = n$ and $\xi_{n+1} = 1$ implies that $H(t) = 0$.

Next, observe that event

$$\left\{R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, S(t) = 1, N(t) = n, \sum_{j=1}^{k}\xi_j = k, \sum_{j=k+1}^{n}\xi_j = 0, \xi_{n+1} = 1\right\}$$

can be rewritten as

$$\begin{aligned}&\left\{R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, \sum_{j=1}^{n+1}M_j + \sum_{j=1}^{k}R_j + \sum_{j=k+1}^{n}H_j \le t,\right.\\ &\left.\sum_{j=1}^{n+1}M_j + \sum_{j=1}^{k}R_j + \sum_{j=k+1}^{n}H_j + R_{n+1} > t, \sum_{j=1}^{k}\xi_j = k, \sum_{j=k+1}^{n}\xi_j = 0, \xi_{n+1} = 1\right\}.\end{aligned}$$

Finally, taking into account that in the case when $\{S(t) = 1\}$ occupation time $R(t) = t - \sum_{j=1}^{n+1}M_j - \sum_{j=k+1}^{n}H_j$ and occupation time $H(t) = \sum_{j=k+1}^{n}H_j$, we get that

$$\begin{aligned}&p_{RH1}(u,v,t)\mathrm{d}u\mathrm{d}v\\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\binom{n}{k}p_1^{k+1}p_2^{n-k}\\ &\quad\times \Pr\left(\sum_{j=1}^{n+1}M_j \in t - \mathrm{d}u - \mathrm{d}v, \sum_{j=k+1}^{n}H_j \in \mathrm{d}v, \sum_{j=1}^{k}R_j \le u, \sum_{j=1}^{k}R_j + R_{n+1} > u\right)\\ &= \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\binom{n}{k}p_1^{k+1}p_2^{n-k}\\ &\quad\times f_M^{(n+1)}(t-u-v)f_H^{(n-k)}(v)\left[F_R^{(k)}(u) - F_R^{(k+1)}(u)\right]\mathrm{d}u\mathrm{d}v.\end{aligned}$$

□

One can verify now that, for instance,

$$p_{R1}(u,t) = \sum_{n=0}^{\infty} p_1^{n+1} f_M^{(n+1)}(t-u)\left[F_R^{(n)}(u) - F_R^{(n+1)}(u)\right] + \int_0^{t-u} p_{RH1}(u,v,t)\mathrm{d}v.$$

Using symmetry between $R(t)$ and $H(t)$, we also can get that

$$p_{RH2}(u,v,t) = \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\binom{n}{k} p_2^{k+1} p_1^{n-k} \times f_M^{(n+1)}(t-u-v) f_R^{(n-k)}(u)\left[F_H^{(k)}(v) - F_H^{(k+1)}(v)\right].$$

Remark 1. *Theorem 1 and Theorem 2 can be generalized to the multiple off states case. We give an example for three off states here. Define iid random variables $\{V_k\}_{k\geq 1}$ with cumulative distribution F_V and probability density function f_V, as the holding times for the 3rd off state (medium repair). Then, $\{\xi_k\}_{k\geq 1}$ will change to iid random variables with $\Pr(\xi_k = R) = p_1$, $\Pr(\xi_k = H) = p_2$, $\Pr(\xi_k = V) = p_3$, and $p_1 + p_2 + p_3 = 1$. Accordingly, the occupation time in the 3rd off-state is $V(t) = \int_0^t 1\{S(s) = 3\}\mathrm{d}s$. By the same technique, we can get the marginal and joint distribution of the occupation times. The only difference is that binomial probabilities are replaced with multinomial ones. For example,*

$$p_{R0}(s,t) = \Pr(R(t) \in \mathrm{d}s, S(t) = 0)/\mathrm{d}s$$

$$= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \sum_{k^*=0}^{n-k} \binom{n}{k} \binom{n-k}{k^*} p_1^k p_2^{k^*} p_3^{n-k-k^*} f_R^{(k)}(s)$$
$$\times \left[F_M^{(n)} * F_H^{(k^*)} * F_V^{(n-k-k^*)}(t-s) - F_M^{(n+1)} * F_H^{(k^*)} * F_V^{(n-k-k^*)}(t-s) \right],$$

and

$$p_{RHV1}(u, v, w; t)$$
$$= \Pr(R(t) \in \mathrm{d}u, H(t) \in \mathrm{d}v, V(t) \in \mathrm{d}w, S(t) = 1)/\mathrm{d}u\mathrm{d}v\mathrm{d}w$$
$$= \sum_{n=0}^{\infty} \sum_{k=0}^{n-1} \sum_{k^*=1}^{n-k-1} \binom{n}{k} \binom{n-k}{k^*} p_1^k p_2^{k^*} p_3^{n-k-k^*} f_M^{(n+1)}(t-v-w-u)$$
$$\times f_H^{(k^*)}(v) f_V^{(n-k-k^*)}(w) \left[F_R^{(k)}(u) - F_R^{(k+1)}(u) \right].$$

All other formulas can be derived similarly.

For the completeness of article, we demonstrate the detailed derivation here. First, let us derive $p_{R0}(s,t) = \Pr(R(t) \in \mathrm{d}s, S(t) = 0)/\mathrm{d}s$*, that is the example for getting marginal distribution. Note that*

$$\Pr(R(t) \in \mathrm{d}s, S(t) = 0) = \sum_{n=1}^{\infty} \Pr\left(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = n\right)$$
$$= \sum_{n=1}^{\infty} \sum_{k=1}^{n} \sum_{k^*=0}^{n-k} \Pr\left(R(t) \in \mathrm{d}s, S(t) = 0, N(t) = n, \sum_{j=1}^{n} 1(\xi_j = R) = k, \sum_{j=1}^{n} 1(\xi_j = H) = k^* \right)$$

where $N(t)$ is defined as Formula 2.1. Then, we have

$$
\begin{aligned}
&\Pr\left(R(t)\in \mathrm{d}s, S(t)=0, N(t)=n, \sum_{j=1}^{n}1(\xi_j=R)=k, \sum_{j=1}^{n}1(\xi_j=H)=k^*\right)\\
=&\Pr\left(R(t)\in \mathrm{d}s, \sum_{j=1}^{n}(M_j+U_j)\le t, \sum_{j=1}^{n}(M_j+U_j)+M_{n+1}>t,\right.\\
&\qquad\left.\sum_{j=1}^{n}1(\xi_j=R)=k, \sum_{j=1}^{n}1(\xi_j=H)=k^*\right)\\
=&\binom{n}{k}\binom{n-k}{k^*}\Pr\left(R(t)\in \mathrm{d}s, \sum_{j=1}^{n}(M_j+U_j)\le t, \sum_{j=1}^{n}(M_j+U_j)+M_{n+1}>t,\right.\\
&\qquad\qquad\sum_{j=1}^{k}1(\xi_j=R)=k, \sum_{j=k+1}^{k+k^*}1(\xi_j=H)=k^*,\\
&\qquad\qquad\left.\sum_{j=k+k^*+1}^{n}1(\xi_j=V)=n-k-k^*\right),
\end{aligned}
$$

where

$$
\begin{aligned}
&\Pr\left(R(t)\in \mathrm{d}s, \sum_{j=1}^{n}(M_j+U_j)\le t, \sum_{j=1}^{n}(M_j+U_j)+M_{n+1}>t,\right.\\
&\qquad\left.\sum_{j=1}^{k}1(\xi_j=R)=k, \sum_{j=k+1}^{k+k^*}1(\xi_j=H)=k^*, \sum_{j=k+k^*+1}^{n}1(\xi_j=V)=n-k-k^*\right)\\
=&p_1^k p_2^{k^*} p_3^{n-k-k^*}\Pr\left(\sum_{j=1}^{k}R_j\in \mathrm{d}s, \sum_{j=1}^{n}M_j+\sum_{j=1}^{k}R_j+\sum_{j=k+1}^{k+k^*}H_j+\sum_{j=k+k^*+1}^{n}V_j\le t,\right.\\
&\qquad\left.\sum_{j=1}^{n+1}M_j+\sum_{j=1}^{k}R_j+\sum_{j=k+1}^{k+k^*}H_j+\sum_{j=k+k^*+1}^{n}V_j>t\right)\\
=&p_1^k p_2^{k^*} p_3^{n-k-k^*}\Pr\left(\sum_{j=1}^{k}R_j\in \mathrm{d}s\right)\left[\Pr\left(\sum_{j=1}^{n}M_j+\sum_{j=k+1}^{k+k^*}H_j+\sum_{j=k+k^*+1}^{n}V_j\le t-s\right)-\right.
\end{aligned}
$$

$$\Pr\left(\sum_{j=1}^{n+1} M_j + \sum_{j=k+1}^{k+k^*} H_j + \sum_{j=k+k^*+1}^{n} V_j \le t-s\right)\Bigg]$$

$$= p_1^k p_2^{k^*} p_3^{n-k-k^*} f_R^{(k)}(s) \Big[F_M^{(n)} * F_H^{(k^*)} * F_V^{(n-k-k^*)}(t-s) -$$

$$F_M^{(n+1)} * F_H^{(k^*)} * F_V^{(n-k-k^*)}(t-s)\Big]\,\mathrm{d}s.$$

Finally, we have

$$p_{R0}(s,t) = \sum_{n=1}^{\infty}\sum_{k=1}^{n}\sum_{k^*=0}^{n-k}\binom{n}{k}\binom{n-k}{k^*} p_1^k p_2^{k^*} p_3^{n-k-k^*} f_R^{(k)}(s)$$

$$\times\left[F_M^{(n)} * F_H^{(k^*)} * F_V^{(n-k-k^*)}(t-s) - F_M^{(n+1)} * F_H^{(k^*)} * F_V^{(n-k-k^*)}(t-s)\right].$$

Next, let us derive $p_{RHV1}(u,v,w;t)$, that is

$$\Pr(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w, S(t)=1)/\mathrm{d}u\mathrm{d}v\mathrm{d}w.$$

This is the example for getting joint distribution. Het us start from

$$\Pr(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w, S(t)=1)$$

$$= \sum_{n=0}^{\infty}\Pr(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w, S(t)=1, N(t)=n)$$

$$= \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\sum_{k^*=1}^{n-k-1}\Pr\Bigg(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w, S(t)=1, N(t)=n,$$

$$\sum_{j=1}^{n} 1(\xi_j = R) = k, \sum_{j=1}^{n} 1(\xi_j = H) = k^*, \xi_{n+1}=1\Bigg)$$

$$=\sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\sum_{k^*=1}^{n-k-1}\Pr\Bigg(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w, S(t)=1, N(t)=n,$$
$$\sum_{j=1}^{k}1(\xi_j=R)=k, \sum_{j=k+1}^{k+k^*}1(\xi_j=H)=k^*, \xi_{n+1}=1\Bigg)\binom{n}{k}\binom{n-k}{k^*},$$

where

$$\Pr\Bigg(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w, S(t)=1, N(t)=n,$$
$$\sum_{j=1}^{k}1(\xi_j=R)=k, \sum_{j=k+1}^{k+k^*}1(\xi_j=H)=k^*, \xi_{n+1}=1\Bigg)$$
$$=\Pr\Bigg(R(t)\in \mathrm{d}u, H(t)\in \mathrm{d}v, V(t)\in \mathrm{d}w,$$
$$\sum_{j=1}^{n+1}M_j+\sum_{j=1}^{k}R_j+\sum_{j=k+1}^{k+k^*}H_j+\sum_{j=k+k^*+1}^{n}V_j\le t,$$
$$\sum_{j=1}^{n+1}M_j+\sum_{j=1}^{k}R_j+\sum_{j=k+1}^{k+k^*}H_j+\sum_{j=k+k^*+1}^{n}V_j+R_{n+1}> t,$$
$$\sum_{j=1}^{k}1(\xi_j=R)=k, \sum_{j=k+1}^{k+k^*}1(\xi_j=H)=k^*, \xi_{n+1}=1\Bigg),$$
$$\textit{note that } R(t)=t-\sum_{j=1}^{n+1}M_j+\sum_{j=k+1}^{k+k^*}H_j+\sum_{j=k+k^*+1}^{n}V_j\in \mathrm{d}u,$$
$$=\Pr\Bigg(\sum_{j=1}^{n+1}M_j\in t-\mathrm{d}v-\mathrm{d}w-\mathrm{d}u, \sum_{j=k+1}^{k+k^*}H_j\in \mathrm{d}v, \sum_{j=k+k^*+1}^{n}V_j\in \mathrm{d}w,$$
$$\sum_{j=1}^{k}R_j\le u, \sum_{j=1}^{k}R_j+R_{n+1}>u\Bigg)p_1^k p_2^{k^*}p_3^{n-k-k^*}$$
$$=p_1^k p_2^{k^*}p_3^{n-k-k^*}f_M^{(n+1)}(t-v-w-u)f_H^{(k^*)}(v)f_V^{(n-k-k^*)}(w)$$
$$\times\Big[F_R^{(k)}(u)-F_R^{(k+1)}(u)\Big]\mathrm{d}u\mathrm{d}v\mathrm{d}w.$$

Finally, we have

$$p_{RHV1}(u,v,w;t) = \sum_{n=0}^{\infty}\sum_{k=0}^{n-1}\sum_{k^*=1}^{n-k-1}\binom{n}{k}\binom{n-k}{k^*}p_1^k p_2^{k^*} p_3^{n-k-k^*} f_M^{(n+1)}(t-v-w-u) \times f_H^{(k^*)}(v) f_V^{(n-k-k^*)}(w)\left[F_R^{(k)}(u) - F_R^{(k+1)}(u)\right].$$

2.4 Special Case: Lévy Distribution

In this section we consider a special case when all the holding times have the Lévy distribution (with location parameter 0 and possibly different scale parameters). The Lévy distribution with scale parameter c^2 has pdf

$$g_c(x) = \frac{c}{\sqrt{2\pi}}\frac{e^{-\frac{c^2}{2x}}}{x^{3/2}}, \quad x > 0,$$

and cdf

$$G_c(x) = \frac{2}{\sqrt{\pi}}\int_{\frac{c}{\sqrt{2x}}}^{\infty} e^{-t^2}\mathrm{d}t, \quad x > 0.$$

The Lévy distribution is heavy-tailed with infinite expectation. The median is given by $0.5c^2\left(\texttt{erfc}^{-1}(0.5)\right)^{-2} \approx 2.198112c^2$, where `erfc` is the complementary error function:

$$\texttt{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^{\infty} e^{-t^2}\,\mathrm{d}t.$$

The parametrization that we use is a bit unusual, but it allows us to shorten our

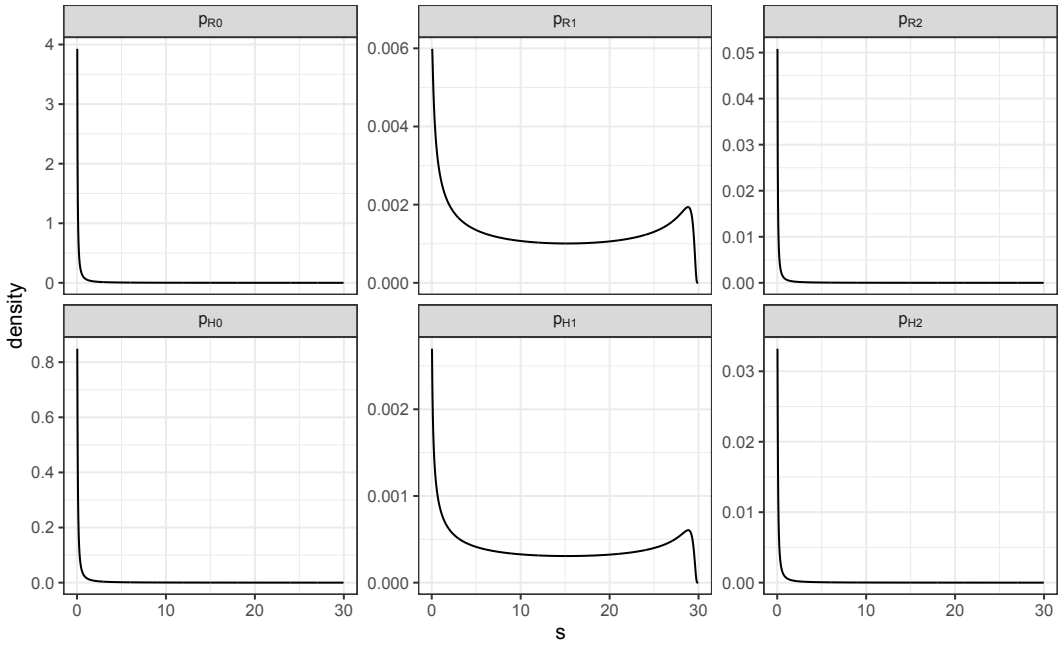

Figure 1: Defective densities $p_{Rj}(s,t)$ and $p_{Hj}(s,t)$, $j \in \{0,1,2\}$, with $t = 30$, $c_M \approx 1.785$, $c_R \approx 0.097$, $c_H \approx 0.275$, and $p_1 = 0.9$.

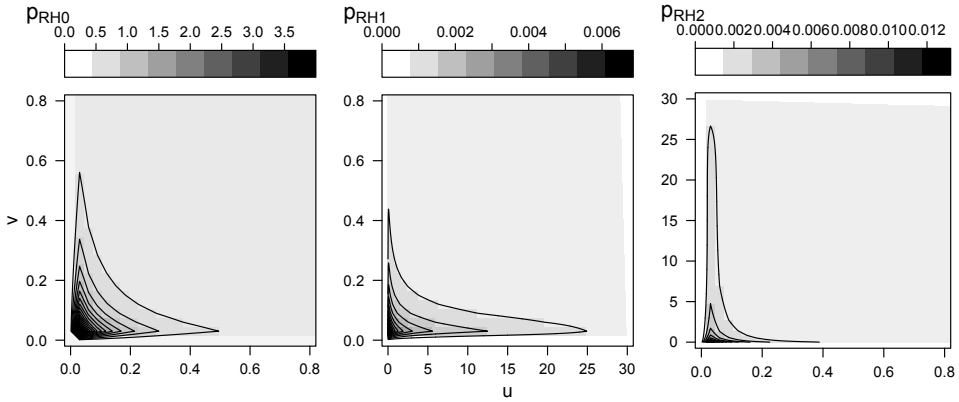

Figure 2: Contour plots of the joint densities $p_{RHj}(u,v,t)$, $j \in \{0,1,2\}$, with $t = 30$, $c_M \approx 1.785$, $c_R \approx 0.097$, $c_H \approx 0.275$, and $p_1 = 0.9$.

notation for convolutions. More specifically, as a member of the family of stable distributions with mobility parameter 1/2, the Lévy distribution is closed under the operation of convolution in the following way:

$$G_{c_1} * G_{c_2}(x) = G_{c_1+c_2}(x).$$

Let c_M^2, c_R^2, and c_H^2 be the scale parameters for states 0, 1 and 2, respectively. Then the formulas from Theorem 1 and Theorem 2 *do not involve* any convolutions. For instance, in this case we have that

$$\begin{aligned} p_{R2}(s,t) = &\sum_{n=1}^{\infty}\sum_{k=1}^{n}\binom{n}{k}p_1^k p_2^{n-k+1} g_{kc_R}(s) \\ &\times \left[G_{(n+1)c_M+(n-k)c_H}(t-s) - G_{(n+1)c_M+(n-k+1)c_H}(t-s)\right]. \end{aligned}$$

The expectations and variances of the off times is given by

$$\begin{aligned} \mathrm{E}(R(t)) &= \sum_{j=0}^{2}\int_0^t s p_{Rj}(s,t)\mathrm{d}s, \quad \mathrm{Var}(R(t)) = \sum_{j=0}^{2}\int_0^t s^2 p_{Rj}(s,t)\mathrm{d}s - [\mathrm{E}(R(t))]^2 \\ \mathrm{E}(H(t)) &= \sum_{j=0}^{2}\int_0^t s p_{Hj}(s,t)\mathrm{d}s, \quad \mathrm{Var}(H(t)) = \sum_{j=0}^{2}\int_0^t s^2 p_{Hj}(s,t)\mathrm{d}s - [\mathrm{E}(H(t))]^2. \end{aligned}$$

Note that the discrete component of the occupation times is not used for these calculations, because the atoms are at 0. The covariance of $R(t)$ and $H(t)$ is given by

$$\mathrm{Cov}(R(t), H(t)) = \sum_{j=0}^{2} \iint_{u,v>0,\, u+v<t} uv p_{RHj}(u, v, t)\mathrm{d}u\mathrm{d}v - \mathrm{E}(R(t))\mathrm{E}(H(t)).$$

As an example we consider a server with median holding time 7 days, 0.5 hour, and 4 hours for the on-state, the short off-state, and the long off-state, respectively. Using one day as the time unit, the model parameters are: $c_M \approx 1.785$, $c_R \approx 0.097$, and $c_H \approx 0.275$. Assume also that the total repair cost is given by

$$C(t) = R(t) + 2H(t),$$

that is, long repairs are twice costlier than the short ones.

Figure 1 presents defective marginal densities $p_{Rj}(s, t)$ and $p_{Hj}(s, t)$ when $t = 30$ days and $p_1 = .9$ (that is, the less serious breakdowns occur 9 times more often). The marginal densities for both occupation times are severely defected when $S(30) = 0$ (the process is in the on-state). As we mentioned above $R(30)$ and $H(30)$ have atoms at $s = 0$:

$$\Pr(R(30) = 0, S(30) = 0) \approx 0.280; \qquad \Pr(H(30) = 0, S(30) = 0) \approx 0.798;$$

$$\Pr(R(30) = 0, S(30) = 1) \approx 0.000; \qquad \Pr(H(30) = 0, S(30) = 1) \approx 0.034;$$

$$\Pr(R(30) = 0, S(30) = 2) \approx 0.004; \qquad \Pr(H(30) = 0, S(30) = 2) \approx 0.000.$$

Figure 2 shows the contour plot of defective joint densities $p_{RHj}(u, v, t)$ with the same parameter setup, where the negative association of the two occupation times is obvious. We also ran a simulation study to numerically confirm the correctness of Theorem 1 and Theorem 2. A total $1,000,000$ realizations of the above on-off process were generated and the empirical results are consistent with the theoretical densities (not shown).

Table 1 summarizes the expectation and variance of the two types of off-state occupation times and the associated repair cost for $t \in \{30, 60\}$ days and $p_1 \in \{0.70, 0.75, \ldots, 0.95, 0.99\}$. As p_1 increases, the expectation and the variance goes up for the less serious breakdown time but goes down fro the more serious breakdown time. The total repair cost has lower expectation and variance for smaller values of p_1. The correlation of the two off-state occupation times is negative, with a magnitude decreasing as p_1 increases. When t is doubled, the expectation of the occupation times and the total repair cost is slightly more than doubled, which is because the process always starts from the on-state with a random holding time.

Table 1: Summaries of off-state occupation times and total repair cost with $c_M \approx 1.785$, $c_R \approx 0.097$, and $c_H \approx 0.275$ as p_1 increases.

t	p_1	$\mathrm{E}(R(t))$	$\mathrm{Var}(R(t))$	$\mathrm{E}(H(t))$	$\mathrm{Var}(H(t))$	$\mathrm{E}(C(t))$	$\mathrm{Var}(C(t))$	$\rho(R(t), H(t))$
30	0.70	0.81	10.82	0.95	13.01	2.71	60.99	−0.040
	0.75	0.87	11.60	0.79	10.95	2.46	53.71	−0.037
	0.80	0.93	12.38	0.64	8.85	2.20	46.32	−0.035
	0.85	0.99	13.15	0.48	6.70	1.95	38.80	−0.031
	0.90	1.05	13.93	0.32	4.51	1.69	31.16	−0.026
	0.95	1.11	14.70	0.16	2.28	1.43	23.38	−0.019
	0.99	1.16	15.32	0.03	0.46	1.23	17.07	−0.009
60	0.70	1.75	48.20	2.08	57.96	5.92	271.58	−0.040
	0.75	1.88	51.68	1.74	48.81	5.37	239.32	−0.038
	0.80	2.02	55.16	1.40	39.46	4.82	206.48	−0.035
	0.85	2.15	58.64	1.05	29.91	4.26	173.04	−0.031
	0.90	2.29	62.13	0.70	20.16	3.70	139.01	−0.026
	0.95	2.42	65.62	0.35	10.19	3.13	104.37	−0.019
	0.99	2.53	68.41	0.07	2.05	2.67	76.21	−0.009

Chapter 3

Brownian Motion Governed by Telegraph Process with Multiple Off States

3.1 Formal Description of MRH Process

Let $S(t)$, $t \geq 0$, be a continuous-time Markov Chain with the state space $\{0, 1, 2\}$ and the transition rate matrix

$$\mathbf{Q} = \begin{pmatrix} -\lambda_0 & \lambda_0 p_1 & \lambda_0 p_2 \\ \lambda_1 & -\lambda_1 & 0 \\ \lambda_2 & 0 & -\lambda_2 \end{pmatrix} \tag{3.1}$$

where $p_1, p_2, \lambda_0, \lambda_1, \lambda_2 > 0$ and $p_1 + p_2 = 1$. The zero entries in the matrix means that state 1 or state 2 do not transit between themselves; only a transition to state 0 is allowed from either of them. In animal movement modeling, the mean duration in state 0, 1, and 2 are, respectively, $1/\lambda_0$, $1/\lambda_1$, and $1/\lambda_2$. We assume that the initial

distribution ν_0 of $S(0)$ is stationary, that is,

$$\nu_0 = \boldsymbol{\pi} = (\pi_0, \pi_1, \pi_2) = \frac{1}{1/\lambda_0 + p_1/\lambda_1 + p_2/\lambda_2} \left(\frac{1}{\lambda_0}, \frac{p_1}{\lambda_1}, \frac{p_2}{\lambda_2} \right). \tag{3.2}$$

Recall that $\boldsymbol{\pi}$ has to satisfy $0 = \boldsymbol{\pi}\mathbf{Q}$ (e.g., Norris, 1998, p.120).

Let $B(t)$ be the standard BM independent of $S(t)$. Then the MRH process is given by

$$X(t) = \sigma \int_0^t 1_{\{S(s)=0\}} \mathrm{d}B(s), \tag{3.3}$$

where $\sigma > 0$ is an infinitesimal standard deviation.

Estimation of the MRH process parameters $\boldsymbol{\theta} = (\lambda_0, \lambda_1, \lambda_2, p_1, \sigma)$ is based on observations at discrete, possibly irregularly spaced time points. The observed data are represented by the vector of observed changes in location

$$\mathbf{X} = \big(X(t_1) - X(0), X(t_2) - X(t_1), \ldots, X(t_n) - X(t_{n-1})\big),$$

where $0 < t_1 < \cdots < t_n$ are the time points of the observations. As mentioned earlier, the difficulty is that the MRH process itself is not Markov. However, the location-state process $\{X(t), S(t)\}$ is Markov. So, our first objective is to derive formulas for transitional probabilities of the location-state process. The key random variable here is

the total time spent in state 0 in the time interval $[0, t]$:

$$M(t) = \int_0^t 1_{\{S(s)=0\}} \mathrm{d}s. \tag{3.4}$$

We also can call this random variable *0-state occupation time* by time t.

A continuous-time Markov Chain can be alternatively described by representing the process $S(t)$ as a combination of a discrete time Markov Chain, holding times, and initial distribution ν. More specifically, let p_{ij} be the probability of switching to state j at the next jump given that we are currently in state i. The matrix

$$\mathbf{P} = (p_{ij}) = \begin{pmatrix} 0 & p_1 & p_2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

is a stochastic matrix, and it is the transition matrix of the embedded (discrete time) Markov Chain of process $S(t)$. The time spent in a particular state i between two consecutive jumps is called the holding time. The holding time has exponential distribution with rate λ_i. For our task this representation (via an embedded Markov Chain and holding times) is a bit more convenient. Note also that in the case of the standard

telegraph process the associated stochastic matrix of the embedded Markov chain is

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Our technique is different from the general approach to the distribution of occupation times in homogeneous finite-state Markov processes (e.g., Sericola, 2000). To develop computationally efficient estimation procedure we exploit the specific structure of our Markov chain. More specifically, a telegraph process can be associated with $S(t)$ if we collapse states 1 and 2 into one state. For this new state the holding time is distributed as a mixture of two exponential distributions. As a consequence, the telegraph process is not Markov. This makes computing the likelihood function for $\mathbf{X}$ challenging, because algorithms like the forward algorithm are not applicable. That is, results for telegraph processes can not be directly employed, because we do need to distinguish states 1 and 2. We use a certain periodicity of the Markov Chain and extend the technique developed in Di Crescenzo (2001) for telegraph processes to obtain the joint distribution of $M(t)$ and $S(t)$. An alternative approach can be developed by extending the method presented in Zacks (2012).

3.2 Joint Distribution of Occupation Time $M(t)$ and $S(t)$

We want to mention that $S(t)$ defined in moving-resting-handling process is a special case of on-off process with multiple off state, with holding times of three states are distributed exponentially. To simulate process $S(t)$ that starts with $S(0) = 0$, we need the following independent sequences of random variables:

1. $\{M_k\}_{k\geq 1}$ are independent identically distributed (iid) random variables with $\text{Exp}(\lambda_0)$ distribution,

2. $\{R_k\}_{k\geq 1}$ are iid random variables with $\text{Exp}(\lambda_1)$,

3. $\{H_k\}_{k\geq 1}$ are iid random variables with $\text{Exp}(\lambda_2)$,

4. $\{\xi_k\}_{k\geq 1}$ are iid random variables with $\Pr(\xi_k = 1) = p_1$ and $\Pr(\xi_k = 0) = p_2$.

Having these sequences defined we can proceed as follows. To generate a particular realization of $S(t)$, first, generate M_1, the time duration the process spends in state 0. Then generate ξ_1 to decide whether it jumps to state 1 or 2. Depending on ξ_1 generate the duration R_1 or H_1. After that, switch back to state 0, and so on.

Let us introduce some auxiliary random variables. Let $U_k = \xi_k R_k + (1 - \xi_k)H_k$, $O_k = M_k + U_k$, and

$$N(t) = \sup\{n \geq 0 : \sum_{k=1}^{n} O_k \leq t\}.$$

Here and everywhere in the text, by convention, a summation over an empty set is 0, for instance, $\sum_{k=1}^{0} O_k = 0$. Random variable $N(t)$ is the number of full cycles O_k by time t.

First, we consider the distribution of occupation time $M(t)$ when $S(t) = 0$. Denote $P_i(\cdot) = \Pr(\cdot|S(0) = i)$, where $i = 0, 1, 2$. With probability 1 the random variable $M(t) \in [0, t]$, and it has an atom at t in the following sense:

$$P_0(M(t) = t, S(t) = 0) = P_0(M(t) = t) = \Pr(M_1 > t) = e^{-\lambda_0 t}.$$

Now, fix $0 < s < t$. Then we have

$$\begin{aligned} P_0(M(t) \in \mathrm{d}s, S(t) = 0) &= \sum_{n=0}^{\infty} P_0(M(t) \in \mathrm{d}s, S(t) = 0, N(t) = n) \\ &= \sum_{n=1}^{\infty} P_0(M(t) \in \mathrm{d}s, S(t) = 0, N(t) = n), \end{aligned}$$

because $S(0) = 0$, $S(t) = 0$ and $N(t) = 0$ implies $M(t) = t$.

Next, for $n \geq 1$ we get that

$$\begin{aligned} & P_0(M(t) \in \mathrm{d}s, S(t) = 0, N(t) = n) \\ &= \Pr\left(\sum_{k=1}^{n} M_k + \sum_{k=1}^{n} U_k \leq t,\ \sum_{k=1}^{n+1} M_k + \sum_{k=1}^{n} U_k > t,\ t - \sum_{k=1}^{n} U_k \in \mathrm{d}s\right) \\ &= \Pr\left(\sum_{k=1}^{n} M_k \leq s,\ \sum_{k=1}^{n+1} M_k > s,\ \sum_{k=1}^{n} U_k \in t - \mathrm{d}s\right) \end{aligned}$$

$$= \Pr\left(\sum_{k=1}^{n} M_k \leq s,\ \sum_{k=1}^{n+1} M_k > s\right) \Pr\left(\sum_{k=1}^{n} U_k \in t - \mathrm{d}s\right)$$
$$= \left[\Pr\left(\sum_{k=1}^{n} M_k \leq s\right) - \Pr\left(\sum_{k=1}^{n+1} M_k \leq s\right)\right] \Pr\left(\sum_{k=1}^{n} U_k \in t - \mathrm{d}s\right).$$

Here we use independence of $\{M_k\}_{k\geq 1}$ and $\{U_k\}_{k\geq 1}$.

The sums $\sum_{k=1}^{n} M_k$ and $\sum_{k=1}^{n+1} M_k$ have gamma distributions, Gamma(n, λ_0) and Gamma$(n+1, \lambda_0)$, respectively. The distribution of $\sum_{k=1}^{n} U_k$ can be expressed in terms of the convolution of gamma distributions. More specifically, by conditioning on $\{\xi_k\}_{1\leq k\leq n}$ one can show that

$$\Pr\left(\sum_{k=1}^{n} U_k \leq s\right) = \sum_{k=0}^{n} \Pr\left(\sum_{j=1}^{k} R_j + \sum_{j=1}^{n-k} H_j \leq s\right) \binom{n}{k} p_1^k p_2^{n-k}.$$

Random variables $\sum_{j=1}^{k} R_j$ and $\sum_{j=1}^{n-k} H_j$ are independent, and they have Gamma(k, λ_1) and Gamma$(n-k, \lambda_2)$ distributions, respectively. For the convolution of gamma distributions, we refer the reader to Mathai (1982), Moschopoulos (1985) and Hu et al. (2020a). The implementations of these methods are publicly available in R package `coga` (Hu et al., 2019).

Next, let us work out the case when $S(t) = 1$. Again, the random variable $M(t) \in [0, t]$, but now it has no atoms. For any $0 < s < t$, we have

$$P_0(M(t) \in \mathrm{d}s, S(t) = 1) = \sum_{n=0}^{\infty} P_0(M(t) \in \mathrm{d}s, S(t) = 1, N(t) = n)$$

Then for $n \geq 0$ we now get that

$$
\begin{aligned}
&P_0(M(t) \in \mathrm{d}s, S(t) = 1, N(t) = n) \\
&= \Pr\left(\sum_{k=1}^{n+1} M_k + \sum_{k=1}^{n} U_k \leq t,\ \sum_{k=1}^{n+1} M_k + \sum_{k=1}^{n} U_k + R_{n+1} > t,\ \sum_{k=1}^{n+1} M_k \in \mathrm{d}s,\ \xi_{n+1} = 1\right) \\
&= \Pr\left(\sum_{k=1}^{n} U_k \leq t - s,\ \sum_{k=1}^{n} U_k + R_{n+1} > t - s,\ \sum_{k=1}^{n+1} M_k \in \mathrm{d}s,\ \xi_{n+1} = 1\right) \\
&= p_1 \Pr\left(\sum_{k=1}^{n} U_k \leq t - s,\ \sum_{k=1}^{n} U_k + R_{n+1} > t - s\right) \Pr\left(\sum_{k=1}^{n+1} M_k \in \mathrm{d}s\right) \\
&= p_1 \left[\Pr\left(\sum_{k=1}^{n} U_k \leq t - s\right) - \Pr\left(\sum_{k=1}^{n} U_k + R_{n+1} \leq t - s\right)\right] \Pr\left(\sum_{k=1}^{n+1} M_k \in \mathrm{d}s\right).
\end{aligned}
$$

Random variable $\sum_{k=1}^{n+1} M_k$ has Gamma$(n+1, \lambda_0)$ distribution. As before,

$$
\Pr\left(\sum_{k=1}^{n} U_k \leq s\right) = \sum_{k=0}^{n} \Pr\left(\sum_{j=1}^{k} R_j + \sum_{j=1}^{n-k} H_j \leq s\right) \binom{n}{k} p_1^k p_2^{n-k},
$$

and

$$
\Pr\left(\sum_{k=1}^{n} U_k + R_{n+1} \leq s\right) = \sum_{k=0}^{n} \Pr\left(\sum_{j=1}^{k+1} R_j + \sum_{j=1}^{n-k} H_j \leq s\right) \binom{n}{k} p_1^k p_2^{n-k}.
$$

The formulas for the joint distribution of $M(t)$ and $S(t)$ given $S(0) = 1$ can be derived in a similar way.

To summarize our findings let us first introduce the following notation:

1. $G(x, \alpha, \beta)$, where $\alpha \geq 0, \beta > 0$, is the cdf of Gamma(α, β) distribution; by convention, Gamma$(0, \beta)$ distribution is the degenerate distribution with atom 1 at

0;

2. $g(x, \alpha, \beta)$, where $\alpha, \beta > 0$, is the pdf of Gamma(α, β) distribution;

3. $F(x, \alpha_1, \beta_1, \alpha_2, \beta_2)$, where $\alpha_1, \alpha_2 \geq 0, \beta_1, \beta_2 > 0$, is the cdf of the convolution of Gamma(α_1, β_1) and Gamma(α_2, β_2); note that, for example, $F(x, 0, \beta_1, \alpha_2, \beta_2) \equiv G(x, \alpha_2, \beta_2)$;

4. $f(x, \alpha_1, \beta_1, \alpha_2, \beta_2)$, where $\beta_1, \beta_2 > 0$, $\alpha_1, \alpha_2 \geq 0$, and $\alpha_1 + \alpha_2 > 0$, is the pdf of $F(x, \alpha_1, \beta_1, \alpha_2, \beta_2)$;

5. $H(x, \alpha_1, \beta_1, \alpha_2, \beta_2) = F(x, \alpha_1, \beta_1, \alpha_2, \beta_2) - F(x, \alpha_1 + 1, \beta_1, \alpha_2, \beta_2)$, where $\beta_1, \beta_2 > 0$, $\alpha_1, \alpha_2 \geq 0$, and $\alpha_1 + \alpha_2 > 0$, is the difference in cdf with parameters only differing by α_1 versus $\alpha_1 + 1$.

Finally, let us denote the (defective) densities of $M(t)$ as

$$p_{ij}(s, t) = P_i(M(t) \in \mathrm{d}s, S(t) = j)/\mathrm{d}s, \tag{3.5}$$

where $t \geq 0$, $0 < s < t$, $i, j = 0, 1, 2$.

Here is the main result of the section.

Theorem 3. *Let $t \geq 0$ and $0 < s < t$. Then*

$$P_0(M(t) = t, S(t) = 0) = e^{-\lambda_0 t}, \tag{3.6}$$

and

$$P_1(M(t)=0, S(t)=1) = e^{-\lambda_1 t}. \tag{3.7}$$

The defective densities are given by

$$p_{00}(s,t) = \sum_{n=1}^{\infty} [G(s,n,\lambda_0) - G(s,n+1,\lambda_0)] \sum_{k=0}^{n} f(t-s,k,\lambda_1,n-k,\lambda_2) \binom{n}{k} p_1^k p_2^{n-k}, \tag{3.8}$$

$$p_{01}(s,t) = \sum_{n=0}^{\infty} p_1 g(s,n+1,\lambda_0) \sum_{k=0}^{n} H(t-s,k,\lambda_1,n-k,\lambda_2) \binom{n}{k} p_1^k p_2^{n-k}, \tag{3.9}$$

$$p_{02}(s,t) = \sum_{n=0}^{\infty} p_2 g(s,n+1,\lambda_0) \sum_{k=0}^{n} H(t-s,k,\lambda_2,n-k,\lambda_1) \binom{n}{k} p_2^k p_1^{n-k}, \tag{3.10}$$

$$p_{10}(s,t) = \sum_{n=0}^{\infty} [G(s,n,\lambda_0) - G(s,n+1,\lambda_0)] \sum_{k=0}^{n} [f(t-s,k+1,\lambda_1,n-k,\lambda_2)] \binom{n}{k} p_1^k p_2^{n-k}, \tag{3.11}$$

$$p_{11}(s,t) = \sum_{n=1}^{\infty} p_1 g(s,n,\lambda_0) \sum_{k=0}^{n-1} H(t-s,k+1,\lambda_1,n-1-k,\lambda_2) \binom{n-1}{k} p_1^k p_2^{n-1-k}, \tag{3.12}$$

and

$$p_{12}(s,t) = \sum_{n=1}^{\infty} p_2 g(s,n,\lambda_0) \sum_{k=0}^{n-1} H(t-s,n-1-k,\lambda_2,k+1,\lambda_1) \binom{n-1}{k} p_1^k p_2^{n-1-k}. \tag{3.13}$$

Note that formula (3.10) can be obtained from (3.9) by interchanging state 1 and state 2. Also, in order to get densities $p_{2j}(s,t)$, $j = 0,1,2$ (that are not listed in Theorem 3) we simply need to interchange state 1 and state 2 in all the formulas for $p_{1j}(s,t)$, $j = 0,1,2$ of Theorem 3. Equations (3.12) and (3.13) look similar, but they are

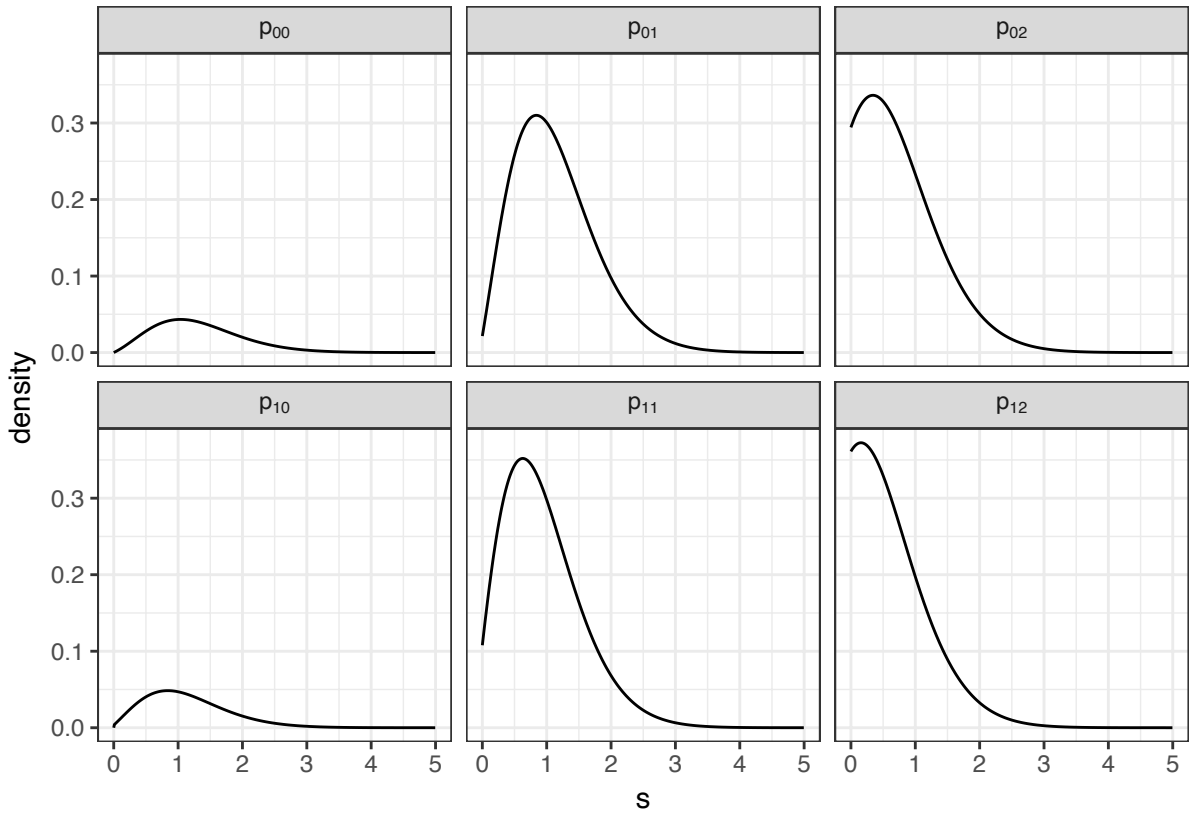

Figure 3: Defective densities $p_{ij}(s,t)$: Theorem 3

significantly different. In (3.12) we start and end in the same off state, but in (3.13) the starting and ending off states are different.

As an example, Figure 3 presents the defective densities $p_{ij}(s,t)$'s for a MRH model with parameters $\lambda_0 = 4$, $\lambda_1 = .5$, $\lambda_2 = .1$, $p_1 = .8$, and $t = 10$. Applications of the formulas in practice depend on how accurately the infinite sums can be implemented. To check the accuracy of the implementation and to verify that our formulas in Theorem 3 are free of errors or typos, we also simulated 1,000,000 realizations of the Markov chain $S(\cdot)$ for each initial state. The empirical densities follow the theoretical ones computed from our implementation extremely closely (not shown).

3.3 Likelihood Estimation with Forward Algorithm

Estimating the parameters of a discretely observed continuous-time process driven by an on-off process (like $S(t)$) is of practical importance but challenging. Besides our work (Yan et al., 2014; Pozdnyakov et al., 2019), only a few authors dealt with the problem. Iacus and Yoshida (2008) and De Gregorio and Iacus (2008) proposed methods that are not based on the true likelihood, such as pseudo-maximum likelihood and moment based estimators. The estimation procedure in De Gregorio and Iacus (2011) was based on a least squares type of method. There are three key differences between our work and these works. First, on top of the on-off process itself we have an additional source of randomness, the Brownian motion. Second, out data are not assumed to be collected at equidistant times. Equidistance is a very restrictive assumption as the animal movement data contain observations that are often irregularly spaced. Finally, asymptotic normality and consistency of our maximum likelihood estimator (MLE) are not addressed in our work. The aforementioned works do give asymptotic results for their estimators but, in order to prove consistency and normality, they need to assume that the time distance between two consecutive observations goes to 0. This assumption is not realistic in the case of animal tracking data. The number of observations (or the time horizon) might be large, but the time intervals between observations are not getting smaller.

Our technique is of an HMM type, but there are two important distinctions from

the classic HMM. First, because of the irregular time spacing, the resulted discrete time Markov process is not time-homogeneous (the transitional probability depends on time). Furthermore, in a classical HMM, given a current hidden state, the corresponding observation is independent of the previous hidden state; this does not hold in our situation. As a consequence, the Baum–Welch algorithm (e.g., Zucchini et al., 2016), which is a special case of the expectation-maximization (EM) algorithm, does not apply in this case. The forward algorithm, however, can still be constructed. The construction relies heavily on formulas for occupation times (Theorem 3). This underscores a practical importance of the numerous theoretical results on occupation times cited in the introduction.

Assume that we observe the MRH process $X(t)$ at times $0 = t_0 < t_1 < \cdots < t_n$. Let $\mathbf{X} = \big(X_1, X_2, \ldots, X_n\big)$, where $X_i = X(t_i) - X(t_{i-1})$, $i = 1, \ldots, n$ are the observed increments of the MRH process. Let $\mathbf{S} = \big(S(0), S(t_1), \ldots, S(t_n)\big)$ be the corresponding states of the continuous-time Markov Chain, and $\Delta_i = t_i - t_{i-1}$, $i = 1, \ldots, n$. To derive the likelihood of the vector of observed increments $\mathbf{X}$, we need a joint distribution of $X(t)$ and $S(t)$ first.

Let us work out the details the formula for $P_0(X(t) \in \mathrm{d}x, S(t) = 0)$. Fix $0 < s < t$. Given $M(t) = s$, random variable $X(t)$ has a normal distribution with mean 0 and variance $\sigma^2 s$, because Markov Chain $S(\cdot)$ and Brownian Motion $B(\cdot)$ are independent processes. Let $\phi(\cdot, \sigma^2)$ denote the pdf of a normal random variable with mean zero and

variance σ^2. Then we get

$$P_0(X(t) \in \mathrm{d}x, S(t) = 0, M(t) \in \mathrm{d}s) = \phi(x, \sigma^2 s) p_{00}(s,t) \mathrm{d}x \mathrm{d}s.$$

Now, recall also that given $S(0) = 0$, random variable $M(t)$ has an atom (with weight $e^{-\lambda_0 t}$ at $s = t$). Therefore, when we integrate s out of the joint distribution of $X(t)$, $S(t)$ and $M(t)$, we get that

$$h_{00}(x,t) = P_0(X(t) \in \mathrm{d}x, S(t) = 0)/\mathrm{d}x = e^{-\lambda_0 t} \phi(x, \sigma^2 t) + \int_0^t \phi(x, \sigma^2 s) p_{00}(s,t) \mathrm{d}s. \tag{3.14}$$

In a similar fashion, one can show that for $i = 1, 2$

$$h_{0i}(x,t) = P_0(X(t) \in \mathrm{d}x, S(t) = i)/\mathrm{d}x = \int_0^t \phi(x, \sigma^2 s) p_{0i}(s,t) \mathrm{d}s. \tag{3.15}$$

When $S(0) = 1$, the distribution of random variable $X(t)$ has an atom at $x = 0$ (if $R_0 > t$, that is, the Markov chain stays in state 1 till time t). Taking this into an account we have the following formulas:

$$h_{1i}(x,t) = P_1(X(t) \in \mathrm{d}x, S(t) = i)/\mathrm{d}x = \int_0^t \phi(x, \sigma^2 s) p_{1i}(s,t) \mathrm{d}s, \text{ if } x \neq 0 \text{ and } i = 0, 1, 2, \tag{3.16}$$

and

$$P_1(X(t) = 0, S(t) = 1) = e^{-\lambda_1 t}.$$

Similarly,

$$h_{2i}(x,t) = P_2(X(t) \in \mathrm{d}x, S(t) = i)/\mathrm{d}x = \int_0^t \phi(x, \sigma^2 s) p_{2i}(s,t)\mathrm{d}s, \text{ if } x \neq 0 \text{ and } i = 0,1,2, \tag{3.17}$$

and

$$P_2(X(t) = 0, S(t) = 2) = e^{-\lambda_2 t}.$$

It is not essential to use one-dimensional Brownian Motion for these derivations but it simplifies our presentation's notation. If one does want to consider a Brownian Motion of d-dimension, then all we need to do is to substitute the one-dimensional normal pdf in formulas (3.14)–(3.17) by the d-dimensional normal density with mean zero and covariance matrix $\sigma^2 I_d$, where I_d is the d-dimensional identity matrix. Of course, in this case x is a vector in the d-dimensional space, not a scalar. In fact, later when we run simulations and analyze real-world data we will use the two-dimensional setup.

Next, the location-state process $\{X(t), S(t)\}$ is Markov, so the likelihood function of $(\mathbf{X}, \mathbf{S})$ is available in closed-form. More specifically, it is given by

$$L(\mathbf{X}, \mathbf{S}, \boldsymbol{\theta}) = \nu(S(0)) \prod_{i=1}^{n} f\big(X_i, S(t_i) | S(t_{i-1}), \Delta_i, \boldsymbol{\theta}\big), \tag{3.18}$$

where

$$f\big(x,u|v,t,\boldsymbol{\theta}\big)=\begin{cases}0 & v\neq u,\ x=0,\\ 0 & v=u=0,\ x=0,\\ e^{-\lambda_1 t} & v=u=1,\ x=0,\\ e^{-\lambda_2 t} & v=u=2,\ x=0,\\ h_{ij}\big(x,t\big) & v=i,\ u=j,\ x\neq 0,\end{cases} \tag{3.19}$$

$x\in\mathbf{R}$, $u,v=0,1,2$, $t>0$, and $\boldsymbol{\theta}=(\lambda_0,\lambda_1,\lambda_2,p_1,\sigma)$.

The distribution of the increments of the MRH process is a mixture of absolutely continuous and discrete distributions. Therefore, in order to construct the likelihood function we have to use the Radon–Nikodym derivative of the probability distribution relative to a dominating measure that includes an atom at $x=0$. That explains the special sets of formulas in the case when $x=0$.

Now, if the state vector S_t is not observed, then obviously the likelihood of the increment vector $\mathbf{X}$ can be computed using

$$L(\mathbf{X},\boldsymbol{\theta})=\sum_{s_0,\ldots,s_n} L(\mathbf{X},(s_0,\ldots,s_n),\boldsymbol{\theta}),$$

where the summation is taken over all possible trajectories of $\mathbf{S}$. However, this formula is not practical since the number of trajectories grows exponentially as sample size $n\to\infty$. This difficulty is addressed with help of the dynamic programming, namely, the forward

algorithm. This algorithm has a linear complexity with respect to n.

The detailed demonstration of normalized forward algorithm is given. First, we need to introduce forward variables:

$$\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta}) = \sum_{s_0,\ldots,s_{k-1}} \nu(s_0) \prod_{i=1}^{k} f\big(X_i, s_i | s_{i-1}, \Delta_i, \boldsymbol{\theta}\big), \tag{3.20}$$

where $\mathbf{X}_k = \big(X_1, X_2, \ldots, X_k\big)$, and $1 \leq k \leq n$. Then one can show that

$$\alpha(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta}). \tag{3.21}$$

That is, for every k we have three forward variables. To get one $k+1$st forward variable we need to calculate three transitional values in (3.19), multiply each kth forward variable by an appropriate transitional value, and finally sum up these three quantities. The bottom line is that the transition from $\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})$ to $\alpha(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta})$ for each k requires a constant (independent of k) number of operations. Since

$$L(\mathbf{X}, \boldsymbol{\theta}) = \sum_{s_n} \alpha(\mathbf{X}_n, s_n, \boldsymbol{\theta}),$$

we get an algorithm that finds $L(\mathbf{X}, \boldsymbol{\theta})$ with computational complexity that is linear with respect to sample size n.

The next step is to modify the forward variables to address the underflow problem.

The problem is that for large k forward variables $\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})$ might be numerically indistinguishable from zero. To resolve this issue the following normalized forward variables are employed:

$$\bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}) = \frac{\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})}{L(\mathbf{X}_k, \boldsymbol{\theta})}, \tag{3.22}$$

where $L(\mathbf{X}_k, \boldsymbol{\theta}) = \sum_{s_k} \alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})$, the likelihood of vector $\mathbf{X}_k$. Then (3.21) immediately implies that the normalized forward variables satisfy the following equation:

$$\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_k, \boldsymbol{\theta})}{L(\mathbf{X}_{k+1}, \boldsymbol{\theta})} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

If for $0 \leq k \leq n-1$ we define

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_{k+1}, \boldsymbol{\theta})}{L(\mathbf{X}_k, \boldsymbol{\theta})},$$

then one can easily verify that

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \sum_{s_{k+1}} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

Here is the normalized version of the forward algorithm.

1. For observed $\mathbf{X}$ and given parameter vector $\boldsymbol{\theta}$, compute $f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big)$ for all possible pairs (s_k, s_{k+1}), $k = 0, \ldots, n-1$.

2. Base case: $\bar{\alpha}(\mathbf{X}_0, s_0, \boldsymbol{\theta}) = \nu(s_0)$, where $s_0 = 0, 1, 2$.

3. Induction: for $s_{k+1} = 0, 1, 2$ compute $\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta})$ using

$$\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \frac{1}{d(\mathbf{X}_{k+1}, \boldsymbol{\theta})} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

and

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \sum_{s_{k+1}} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

4. Termination: $\log L(\mathbf{X}, \boldsymbol{\theta}) = \sum_{k=1}^{n} \log d(\mathbf{X}_k, \boldsymbol{\theta})$.

The forward algorithm can be easily adapted to a situation when some states are completely observed or partially observed. For example, accelerometer data might be used to infer when an animal is moving or not, and direct inspection of a kill-site can confirm handling. If state s_k is known, then first calculate three kth forward variables as usual. Next, set the two forward variables with unobservable states to zero. After that just continue the forward algorithm in the normal fashion until the next location where additional information on the state is available. If at kth location only one state is excluded, then we have to set only one forward variable to zero.

The MLE can be obtained by maximizing the likelihood function facilitated with the forward algorithm. For the MRH model, however, finding the MLE is extremely computing intensive. First, evaluation of the terms in Theorem 3 involves infinite series that are computationally demanding. Second, evaluation of the terms in the likelihood is also very expensive as functions in (3.19) are numerical integrals of $p_{ij}(s, t)$. In our

investigation, evaluating the likelihood once for a dataset of $n = 200$ takes about 30 minutes on a computer with 3.4 GHz CPU, which is why we could only afford small simulation studies in Section 3.5.

The asymptotic properties the MLE are open questions. Cappé et al. (2005) demonstrates that under certain regularity conditions the MLEs of standard HMM parameters are consistent, asymptotically normal, and asymptotically efficient. As pointed out in Zucchini et al. (2016), however, those conditions do not hold for many practical models, and the asymptotic normality is achieved when the sample size is very large. Similar to what has been in practice, we conjecture that the asymptotic variance matrix of the MLE can be estimated by inverting the Fisher information matrix. Specifically, we numerically evaluate the observed Fisher information matrix, that is, the Hessian matrix or second derivative of the negative log-likelihood function with respect to the parameters, by the finite difference method at the MLE and use its inverse as the variance estimator. Inverting the observed Fisher information is known to be better than inverting the expected Fisher information, even when the latter can be evaluated, in approximating the variance of the MLE (Efron and Hinkley, 1978). Alternatively, the variance matrix of the MLE can be estimated by parametric bootstrap (Efron and Tibshirani, 1994), where each bootstrap sample is generated on the same time grid as the observed data using the fitted parameters. In Section 3.5 we will demonstrate (via simulations) in a one-parameter situation that the average of standard error based on the Fisher information are consistent with the empirical one.

3.4 Prediction of Hidden States: Marginal Probability and Viterbi Path

Another interesting and practical question that can also be addressed by dynamic programming is prediction of hidden states. The prediction question can be formulated in different ways. We consider here two problems.

3.4.1 Marginal probability

First, given parameter set $\boldsymbol{\theta}$ and observations $\mathbf{X}$, what is $\Pr(S(t_k) = i|\mathbf{X})$, the conditional probability that the kth state is equal to i ($i = 0, 1, 2$)? As mentioned near the end of the previous section,

$$L(\mathbf{X}, k, i, \boldsymbol{\theta}) = \sum_{s_0,\ldots,s_n;s_k=i} L(\mathbf{X}, (s_0, \ldots, s_n), \boldsymbol{\theta})$$

can be efficiently computed by the forward algorithm, and then

$$\Pr(S(t_k) = i|\mathbf{X}) = L(\mathbf{X}, k, i, \boldsymbol{\theta})/L(\mathbf{X}, \boldsymbol{\theta})$$

gives us the answer. Nonetheless, if we need a prediction of another state, the forward algorithm must be run again.

This can be avoided by complementing the forward variables with backward ones.

More specifically, in addition to the forward variables given by

$$\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta}) = \sum_{s_0,\ldots,s_{k-1}} \nu(s_0) \prod_{i=1}^{k} f\big(X_i, s_i | s_{i-1}, \Delta_i, \boldsymbol{\theta}\big), \tag{3.23}$$

where $\mathbf{X}_k = \big(X_1, X_2, \ldots, X_k\big)$, and $1 \leq k \leq n$, we also introduce the forward variables:

$$\beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}) = \sum_{s_{k+1},\ldots,s_n} \prod_{i=k+1}^{n} f\big(X_i, s_i | s_{i-1}, \Delta_i, \boldsymbol{\theta}\big), \tag{3.24}$$

where $\tilde{\mathbf{X}}_k = \big(X_{k+1}, X_{k+2}, \ldots, X_n\big)$, and $1 \leq k < n$. Note that the forward variable is the likelihood of observing $\mathbf{X}_k$ *and* $S(t_k) = s_k$, and the backward variable is the likelihood of observing $\bar{\mathbf{X}}_k$ *given* that $S(t_k) = s_k$.

Then one can show that that the forward and backward variables satisfy the following recursive formulas:

$$\alpha(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta}), \tag{3.25}$$

and

$$\beta(\tilde{\mathbf{X}}_{k-1}, s_{k-1}, \boldsymbol{\theta}) = \sum_{s_k} f\big(X_k, s_k | s_{k-1}, \Delta_{k-1}, \boldsymbol{\theta}\big) \beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}). \tag{3.26}$$

The key observation is that the product of forward and backward variables is the likelihood of observing $(\mathbf{X}, S(t_k) = i)$. More specifically, we have that

$$L(\mathbf{X}, k, i, \boldsymbol{\theta}) = \alpha(\mathbf{X}_k, i, \boldsymbol{\theta}) \beta(\tilde{\mathbf{X}}_k, i, \boldsymbol{\theta}).$$

But we are not done yet. The problem is that for large k the forward and backward variables might be numerically indistinguishable from zero. To address the underflow issue we need to introduce normalized forward and backward variables. Specifically, the normalized forward variables are defined by

$$\bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}) = \frac{\alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})}{L(\mathbf{X}_k, \boldsymbol{\theta})}, \tag{3.27}$$

where $L(\mathbf{X}_k, \boldsymbol{\theta}) = \sum_{s_k} \alpha(\mathbf{X}_k, s_k, \boldsymbol{\theta})$, the likelihood of vector $\mathbf{X}_k$. Then (3.21) immediately implies that the normalized forward variables satisfy the following equation:

$$\bar{\alpha}(\mathbf{X}_{k+1}, s_{k+1}, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_k, \boldsymbol{\theta})}{L(\mathbf{X}_{k+1}, \boldsymbol{\theta})} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

If for $0 \leq k \leq n-1$ we define

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_{k+1}, \boldsymbol{\theta})}{L(\mathbf{X}_k, \boldsymbol{\theta})},$$

then one can easily verify that

$$d(\mathbf{X}_{k+1}, \boldsymbol{\theta}) = \sum_{s_{k+1}} \sum_{s_k} f\big(X_{k+1}, s_{k+1} | s_k, \Delta_{k+1}, \boldsymbol{\theta}\big) \bar{\alpha}(\mathbf{X}_k, s_k, \boldsymbol{\theta}).$$

Now, the normalized backward variables are defined by

$$\bar{\beta}(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}) = \frac{L(\mathbf{X}_k, \boldsymbol{\theta})}{L(\mathbf{X}, \boldsymbol{\theta})} \beta(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}). \tag{3.28}$$

Equation (3.26) immediately gives us that

$$\bar{\beta}(\tilde{\mathbf{X}}_{k-1}, s_{k-1}, \boldsymbol{\theta}) = \frac{1}{d(\mathbf{X}_k, \boldsymbol{\theta})} \sum_{s_k} f\big(X_k, s_k | s_{k-1}, \Delta_{k-1}, \boldsymbol{\theta}\big) \bar{\beta}(\tilde{\mathbf{X}}_k, s_k, \boldsymbol{\theta}). \tag{3.29}$$

Here is the procedure for finding $\Pr(S(t_k) = i|\mathbf{X})$. First run the forward algorithm and store all the normalized forward variables, all $d(\mathbf{X}_k, \boldsymbol{\theta})$, and all $f\big(X_k, s_k|s_{k-1}, \Delta_{k-1}, \boldsymbol{\theta}\big)$. Then by employing recursive formula (3.29) (with starting values $\bar{\beta}(\tilde{\mathbf{X}}_n, s_n, \boldsymbol{\theta}) = 1$), calculate and store the normalized backward variables $\beta(\bar{\mathbf{X}}_k, s_k, \boldsymbol{\theta})$. Now, one can easily verify that

$$\Pr(S(t_k) = i|\mathbf{X}) = \bar{\alpha}(\mathbf{X}_k, i, \boldsymbol{\theta}) \bar{\beta}(\tilde{\mathbf{X}}_k, i, \boldsymbol{\theta}).$$

Note that, given observations $\mathbf{X}$ and a set of parameters $\boldsymbol{\theta}$, we need to run both forward and backward algorithms just once.

3.4.2 Viterbi path

The second prediction problem is finding the so-called Viterbi path (Viterbi, 2006) — the most likely sequence of hidden states given observed $\mathbf{X}$ and a fixed parameter set $\boldsymbol{\theta}$. The naive approach is to calculate the likelihood of every path, and then choose the one

with the highest value. The complexity of this procedure is exponential with respect to sample size n. The Viterbi algorithm is a dynamic programming algorithm and is linear in n.

The key idea is to introduce the most likely paths from time 0 to time t_k that end in states $i = 0, 1, 2$, and then by using Markov property to establish a (recursive) relationship between two consecutive likelihoods. More specifically, assume that $(s_0^i, s_1^i, \ldots, s_{k-1}^i, i)$ is the most likely path that ends in the state i given observed vector $\mathbf{X}_k$. Let $L_V(k, i)$ be the likelihood of this path, that is,

$$L_V(k, i) = \Pr(S(0) = s_0^i, S(t_1) = s_1^i, \ldots, S(t_{k-1}) = s_{k-1}^i, S(t_k) = i | \mathbf{X}_k).$$

Then

$$L_V(k+1, i) = \max_{j=0,1,2} \Big\{ L_V(k, j) f\big(X_{k+1}, i | j, \Delta_{k+1}, \boldsymbol{\theta}\big) \Big\},$$

and, respectively, the argument of this maximum gives us the next state of the Viterbi path s_k^i. Once we have all 3 paths $(s_0^i, s_1^i, \ldots, s_{n-1}^i, i)$, $i = 0, 1, 2$, we just choose the one with the highest likelihood.

3.4.3 An illustration

For illustration, consider state prediction based on the marginal probability. We used 1000 realizations of an MRH process with the following parameters: $\lambda_0 = 0.3$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\sigma = 10$, $p_1 = 0.5$, and $S(0) = 0$. The time grid was $(0, 10, 20, \ldots, 2000)$.

Table 2: Predicted probabilities and relative frequency of the states at $t_{100} = 1000$ based on 1000 replicates of an MRH process with parameters $\lambda_0 = 0.3$, $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\sigma = 10$, $p_1 = 0.5$, and $S(0) = 0$ on time grid $(0, 10, 20, \ldots, 2000)$.

	$S(t_{100}) = 0$	$S(t_{100}) = 1$	$S(t_{100}) = 2$
	True state		
Stationary proportion ($\nu(\cdot)$)	0.308	0.231	0.462
Empirical proportion	0.327	0.232	0.441
$\Pr(S(t_{100}) = 0\|\mathbf{X})$	0.484	0.314	0.169
$\Pr(S(t_{100}) = 1\|\mathbf{X})$	0.279	0.286	0.234
$\Pr(S(t_{100}) = 2\|\mathbf{X})$	0.237	0.400	0.597

Table 2 summarizes the marginal probabilities of the states given the observed data in comparison with the stationary probabilities and the empirical relative frequencies. Note that even though each trajectory was started from a moving state, the empirical distribution of the state process at $t_{100} = 1000$ is very close to the stationary one. This is not surprising because by t_{100} the Markov chain has enough time to enter into a stationary stage. The predicted marginal probabilities of the states at t_{100}, however, are noticeably different depending on what the true states are. For example, the average marginal probabilities $\Pr(S(t_{100}) = 0|\mathbf{X})$ evaluated for those trajectories for which $S(t_{100})$ was actually 0 is 0.484, which is much higher then the stationary probability 0.308. This excess in probability is different for different states, and it looks like it follows a very complicated pattern that depends on model parameters and the time distance between observations.

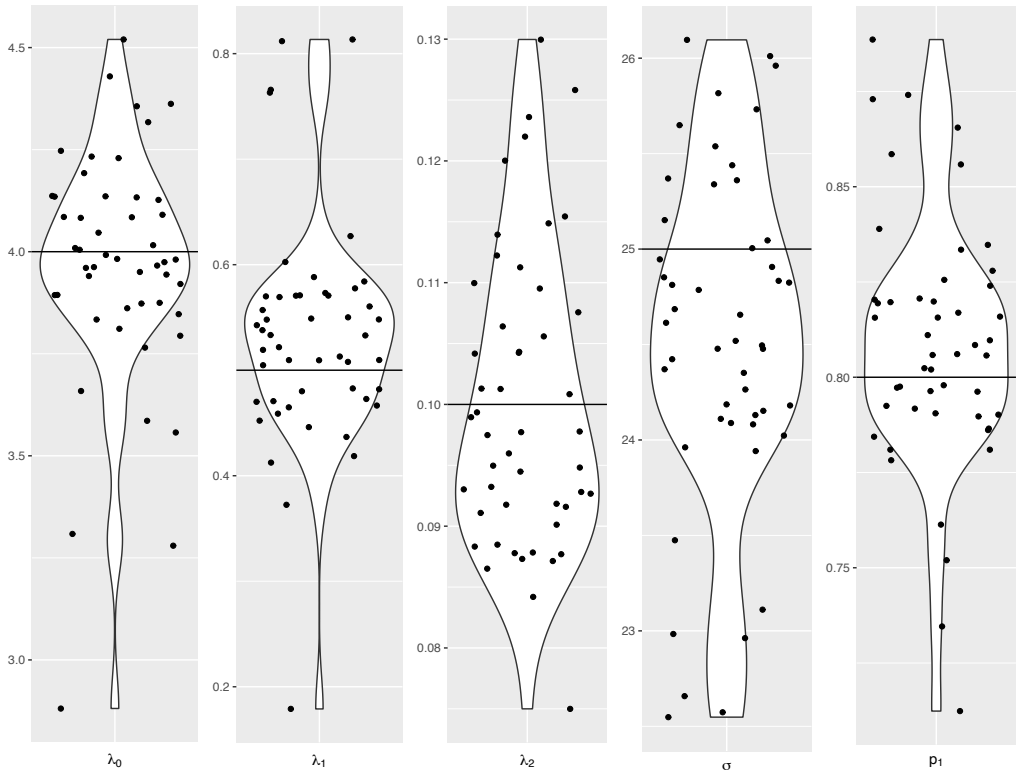

Figure 4: Violin plots of the maximum likelihood estimates from 49 replicates using the forward algorithm. The horizontal bar in each panel is the true parameter value.

3.5 Numerical Studies

We ran a small simulation to demonstrate that the MLE from the forward algorithm successfully recovers the model parameters. The true parameter values were set to be $\lambda_0 = 4$, $\lambda_1 = 0.5$, $\lambda_2 = 0.1$, $p_1 = 0.8$, and $\sigma = 25$. The simulation was small because the computation of the maximum likelihood estimator is very demanding, even with the help of a large beowulf cluster of over 500 CPUs. We generated 49 two-dimensional datasets on a time grid from 0 to 4000, with increment 20, so the resulting series $\mathbf{X}$ is of length 200.

Figure 4 presents the violin plots of the likelihood estimates of the 49 replicates in comparison to the true values of the five parameters. Violin plots are similar to box plots with a rotated kernel density plot on each side, which show more information about the data than box plots. The horizontal bars in the panels are the true parameter values. For each parameter, the true value lies in the bulk part of the violin plot, indicating that the true parameters are recovered well by the likelihood estimates in this small scale simulation study.

A separate study focusing on the standard error of the estimator was conducted. In this study, we only estimate p_1 with all other parameters fixed to make the task computationally feasible within a reasonable period of time. The parameters were set the same as in the last study. We generated 100 two-dimensional datasets, each of which was generated on a time grid from 0 to 400 with equal increment 2. The sampling frequency in this study is much higher than in the last study to make the estimation task easier for computational feasibility. For each datasets, standard errors of the MLE were obtained from both Fisher information and parametric bootstrap. In the parametric bootstrap procedure, bootstrap sample size was set to be 25, and time grid remained the same as that in the observed data. Over the 100 replicates, the averaged estimate for p_1 is 0.083 and the empirical standard error is 0.096. The average standard errors from the Fisher information and parametric bootstrap are, respectively, 0.088 and 0.095. The Fisher information approach under-estimates the standard error; a larger sample size is needed for better agreement. On the other hand, the bootstrap approach estimates

the standard error very well with steep computing cost. This study, although limited in scope, suggests that inferences based on Fisher information or parametric bootstrap are possible. A fuller investigation merits future research.

We next applied the proposed model to the data from the same mountain lion analyzed by Yan et al. (2014) and Pozdnyakov et al. (2019). This mount lion was a mature female in the Gros Ventre Mountain Range near Jackson Wyoming tracked with a GPS collar from 2009 to 2012. The collar was designed to collect a fix every 8 hours but the actual sampling times were irregular with sampling intervals having standard deviation 5.37 hours, ranging from 0.5 hours to 44 hours. Mountain lions behave differently in the summer and in the winter, so we focused on the summer of 2012, a total of 398 observations spanning from June 1 to August 31. This makes our results not directly comparable to existing analyses (Yan et al., 2014; Pozdnyakov et al., 2019). Field personnel determined that some of the sites were places where the mountain lion consumed a prey item. She typically remained within 250 m of a kill site while it was considered to be "handling", which is different from shorter, resting periods. To allow for GPS measurement error, we rounded the locations to the nearest 100 meters.

Table 3 summarizes the MLEs fitted from the MRH process with standard errors obtained from inverting the observed Fisher information matrix. The estimates of the rates of the exponential holding times suggest that, on average, the mountain lions stays for 0.119, 0.755, and 7.692 hours in the moving, resting, and handling states. The mobility parameter estimate means that, if the mountain lion moves without stopping for

Table 3: Parameter estimates and model selection criteria from the mountain lion data analysis with the MRH process and the MR process. The standard errors were obtained from inverting the observed Fisher information matrix.

Parameters	MRH		MR	
	estimate	standard error	estimate	standard error
λ_0	8.402	2.271	3.016	0.320
λ_1	1.324	0.250	0.207	0.013
λ_2	0.130	0.024	—	—
σ	1.563	0.077	1.305	0.064
p	0.685	0.159	—	—
log-likelihood	−982.6		−1021.5	
AIC	1975.2		2049.0	
BIC	1995.1		2061.0	

one hour, the average deviation from the initial position in terms of northing and easting values is 1.563 km. When she stopped moving, she went into resting with probability 0.685 and handling with probability 0.315, respectively. Also reported in the table are the results from fitting the MR process (Yan et al., 2014; Pozdnyakov et al., 2019). Because there is no handling state, the average durations in both moving and resting are estimated longer at 0.332 and 4.831 hours, respectively. The mobility parameter estimate is 17% lower than that in the MRH model, because the animal was assumed to be moving longer. The MRH model has a log-likelihood 139.9 higher than the MR model, which is well worth the two additional parameters. Based on either the Akaike Information Criterion (AIC) or the Bayesian Information Criterion (BIC), the MRH model is very strongly preferred to the MR model. We also fitted the BBMM (Horne et al., 2007; Pozdnyakov et al., 2014) with the original, non-rounded data and the GPS measurement error standard deviation fixed at 0.02km. The BM mobility parameter

estimate is much lower, 0.553km/hour$^{1/2}$, about 35% of that in the MRH model, as the animal was assumed to be always moving.

Chapter 4

Moving-Resting Process with Measurement Error

4.1 Moving-Resting Process

As we mentioned before, the recently proposed moving-resting (MR) process (Yan et al., 2014) is a promised model for animal movement. The MR process is a Brownian motion governed by a telegraph or on-off process (e.g., Zacks, 2004). Specifically, it allows an animal to alternate between a moving state, during which it moves in a Brownian motion (BM), and a resting state, during which it remains motionless. The switch between the two states is characterized governed by a telegraph process, where the holding time (or duration) of each state is assumed to follow an exponential distribution. The memory-less holding time makes the underlying state process a continuous time Markov Chain. As a consequence, the MR process can be analyzed with the help of hidden Markov model (HMM) tools (Cappé et al., 2005). The MR process is a first step towards more realistic animal movement modeling with discretely observed telemetry data where the

trajectories contain evident motionless segments. Implementation of likelihood based inferences for the MR process based on dynamic programming (Pozdnyakov et al., 2019) is publicly available in an R package `smam` (Hu et al., 2020b).

The moving-resting process is a Brownian motion with an infinitesimal variance that is governed by an alternating renewal process with two different holding times. Let random variables $\{M_i\}_{i\geq 1}$ and $\{R_i\}_{i\geq 1}$ be independent and identically distributed random variables that have exponential distributions with rate parameters λ_1 and λ_0, respectively. These are the holding times. There are two possible alternating sequences of the holding times, $(M_1, R_1, M_2, R_2, \dots)$ or $(R_1, M_1, R_2, M_2, \dots)$. Which one will be used for a particular realization is determined by an initial distribution. The continuous time state process, $S(t)$, $t \geq 0$, takes only two values, 0 and 1, and it is defined as follows. For sequence $(M_1, R_1, M_2, R_2, \dots)$, if there exists $k \geq 0$ such that

$$\sum_{j=1}^{k}(M_j + R_j) < t \text{ but } \sum_{j=1}^{k}(M_j + R_j) + M_k \geq t$$

then $S(t) = 1$, otherwise, $S(t) = 0$. For sequence $(R_1, M_1, R_2, M_2, \dots)$, if there exists $k \geq 0$ such that

$$\sum_{j=1}^{k}(R_j + M_j) < t \text{ but } \sum_{j=1}^{k}(R_j + M_j) + R_k \geq t$$

then $S(t) = 0$, otherwise, $S(t) = 1$. It is well-know that the state process is stationary,

if the initial probability of $\{S(0) = 1\}$ is set as

$$p_1 = \frac{\lambda_0}{\lambda_0 + \lambda_1},$$

and the initial probability of $\{S(0) = 0\}$ is set as $p_0 = 1 - p_1$.

The moving-resting process $X(t)$, $t \geq 0$, can be defined with help the following stochastic differential equation

$$\mathrm{d}X(t) = \begin{cases} \sigma \mathrm{d}B(t) & \text{if } S(t) = 1, \\ 0 & \text{if } S(t) = 0, \end{cases}$$

where $B(t)$ is the standard Brownian motion, and σ is a volatility parameter. It is important to note moving-resting process $\{X(t)\}_{t \geq 0}$ itself is not Markov, but the location-state process $\{X(t), S(t)\}$ is Markov with stationary increments.

The moving-resting process has been systematically and comprehensively studied in Yan et al. (2014) and Pozdnyakov et al. (2019). One of the key elements is the distribution of occupation times. More specifically, the total time spent in moving state by time t is given by

$$M(t) = \int_0^t S(s)\mathrm{d}s.$$

Respectively, $R(t) = t - M(t)$ is the total time spent in the resting state. Let $\mathbf{P}_i(\cdot)$ be the conditional probability $\mathbf{P}(\cdot|S(0) = i)$. Zacks (2004) derived computationally efficient

formulas for the following (defective) densities

$$p^*_{11}(w,t)\mathrm{d}w = \mathbf{P}_1(M(t) \in \mathrm{d}w, S(t) = 1),$$

$$p^*_{10}(w,t)\mathrm{d}w = \mathbf{P}_1(M(t) \in \mathrm{d}w, S(t) = 0),$$

$$p^*_{01}(w,t)\mathrm{d}w = \mathbf{P}_0(R(t) \in \mathrm{d}w, S(t) = 1),$$

$$p^*_{00}(w,t)\mathrm{d}w = \mathbf{P}_0(R(t) \in \mathrm{d}w, S(t) = 0),$$

where $0 < w < t$.

Having this at hand, one can derive the marginal distribution of the increment $X(t) - X(0)$. Without loss of generality, let $X(0)$ to be 0, and $X(t)$ becomes the increment from time 0 to time t. Then, the joint distribution of the increment $X(t)$ and $S(t)$ is given by

$$\mathbf{P}_1(X(t) \in \mathrm{d}x, S(t) = 1) = h^*_{11}(x,t)\mathrm{d}x,$$

$$\mathbf{P}_1(X(t) \in \mathrm{d}x, S(t) = 0) = h^*_{10}(x,t)\mathrm{d}x,$$

$$\mathbf{P}_0(X(t) \in \mathrm{d}x, S(t) = 0) = h^*_{00}(x,t)\mathrm{d}x + e^{-\lambda_0 t}\delta_0(x),$$

$$\mathbf{P}_0(X(t) \in \mathrm{d}x, S(t) = 1) = h^*_{01}(x,t)\mathrm{d}x,$$

where $h_{ij}(x,t)$ are functions derived in Yan et al. (2014) and given by

$$h^*_{11}(x,t) = e^{-\lambda_1 t}\phi(x;\sigma^2 t) + \int_0^t \phi(x;\sigma^2 w)p^*_{11}(w,t)\mathrm{d}w,$$

$$h^*_{10}(x,t) = \int_0^t \phi(x;\sigma^2 w) p^*_{10}(w,t)\mathrm{d}w,$$
$$h^*_{00}(x,t) = \int_0^t \phi(x;\sigma^2 (t-w)) p^*_{00}(w,t)\mathrm{d}w,$$
$$h^*_{01}(x,t) = \int_0^t \phi(x;\sigma^2 (t-w)) p^*_{01}(w,t)\mathrm{d}w.$$

Here $\phi(\cdot;\sigma^2)$ is the density function of normal distribution $N(0,\sigma^2)$, and $\delta_0(x)$ is the delta function with an atom at 0, $x \in \mathbb{R}$, and $t \geq 0$. The parameter estimator procedure based on a composite likelihood was discussed in Yan et al. (2014). The maximum likelihood estimation algorithm based on dynamic programming was developed in Pozdnyakov et al. (2019).

In case of real world data sets, however, we may never observe exact values of $X(t)$, but $X(t)$ with added measurement errors. Ignoring the added noise (as it can be done in case of the BBME) does not work even when the noise is relatively small. Addressing this problem via rounding can help. But it is not a trivial task to come up with an appropriate choice for rounding accuracy. Figure 5 (up left) shows the easting/northing coordinates in Universal Transverse Mercator (UTM) of a female lion in 2012 in the Gros Ventre mountain range, Wyoming. The patterns of resting — places where both lines are flat — and moving are readily apparent, which can hardly be captured by any existing model that assumes perpetual movements. The other three panels of Figure 5 show two coordinates of simulated path from an MR process without noise and with noise of two different sizes. The pattern is very similar to that in the left panel for the

female mountain lion. Another less apparent but interesting characteristic of the MR model in comparison to the BM is that the observed increments in easting and northing are dependent. Indeed, if we have a small amount of movement in one direction, it might mean that the animal rested during this time, and, as a result, the corresponding movement in the other direction would likely be small, too.

The significant impact of introducing measurement errors on the estimation of moving-resting process parameters can be shown by simulations. Let us consider a moving-resting process, with parameter $\lambda_1 = 1/hr.$, $\lambda_0 = 0.5/hr.$, $\sigma = 1km/hr.^{1/2}$. We assume measurement error is produced by independent Gaussian white noise. The standard deviation of noise was set as $0.05km$ (1/20 of volatility,) and $0.01km$ (1/100 of volatility) in our simulations. The length of the time intervals between consecutive observations is 5. The number of observations is 200. A traditional approach of handling measurement error is rounding the original observation before performing parameter estimation. The maximum likelihood estimates (based on moving-resting process model) are calculated for randomly generated moving-resting process without measurement errors and with measurement errors. For the series with error, we round the data with different levels of accuracy. The simulation results are recorded in Table 4 as mean and standard deviation of estimators from 100 replications. The point estimation procedure performance is acceptable when series are not contaminated with measurement error. However, when we introduce measurement error into our model, its performance becomes unsatisfactory. Moreover, rounding to 10 meters, 50 meters and 100 meters does not help much.

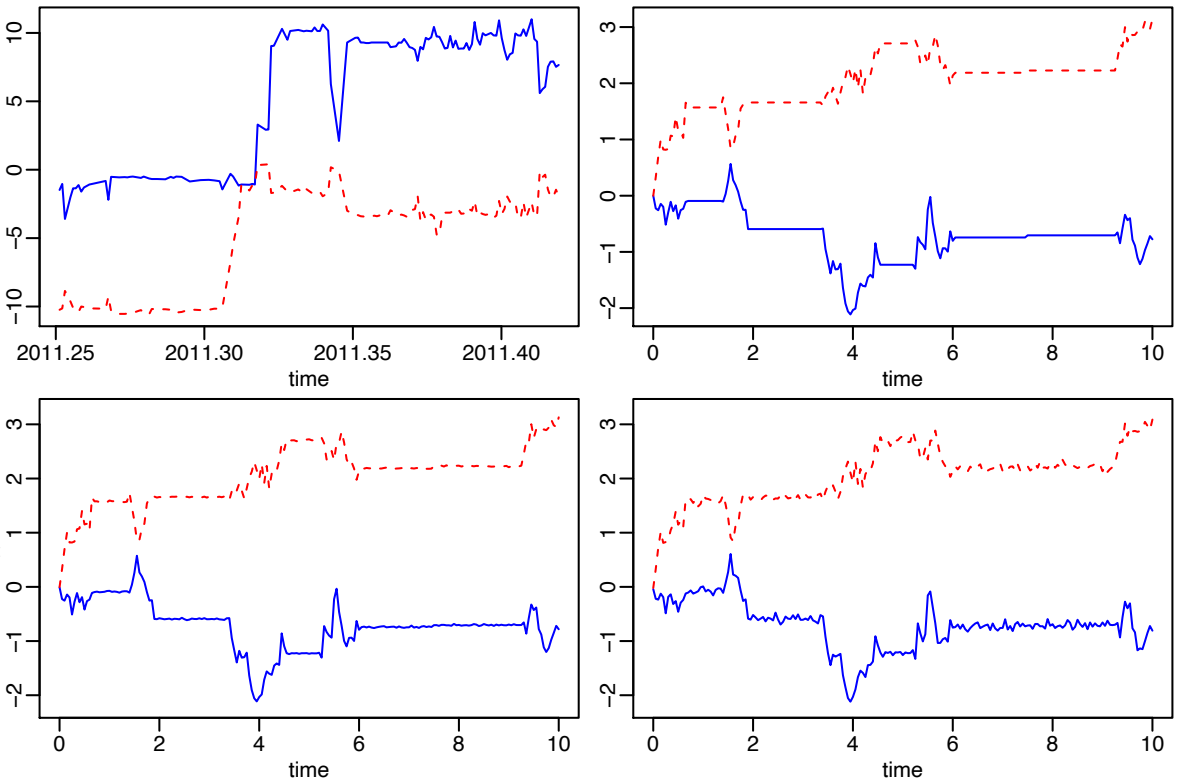

Figure 5: *Up left*: Actual coordinates of a female mountain lion in a two-month period in 2012 in the Gros Ventre mountain range, Wyoming, with most observations separated by 8 hours. The x-axis is time in years. The y-axis is departure from the starting point. The solid blue line is UTM easting (km) and the dashed red line is UTM northing (km). *Up right*: Coordinates of a realization from a two-dimensional MR process. The two coordinates are dependent because the straight line segments representing resting periods are shared. *Bottom left*: Coordinates of the same realization as up right panel after adding Gaussian noise with standard deviation 0.01. *Bottom right*: Coordinates of the same realization as up right panel after adding Gaussian noise with standard deviation 0.05.

Table 4: Influence of measurement error on moving-resting process parameter estimation. The true parameter of moving-resting process is $\lambda_1 = 1/hr$, $\lambda_0 = 0.5/hr$, $\sigma = 1km$. The measurement error is set as Gaussian noise with standard deviation 0.05 and 0.01. The length of the time intervals between consecutive observations is 5. The number of observations is 200. The number of replication is 100. The mean and empirical standard deviation of estimators under different setups are recorded.

Gaussian noise	Rounding	$\hat{\lambda_1}$		$\hat{\lambda_0}$		$\hat{\sigma}$	
s.d. (km)	(km)	mean	s.d.	mean	s.d.	mean	s.d
—	—	1.58	2.29	0.50	0.08	1.08	0.38
0.05	—	21.16	12.21	0.89	0.08	2.73	0.75
	0.01	21.58	9.92	0.89	0.08	2.78	0.68
	0.05	15.08	10.00	0.85	0.08	2.34	0.76
	0.10	9.72	7.70	0.75	0.10	1.98	0.74
0.01	—	22.00	10.82	0.83	0.07	2.89	0.75
	0.01	17.61	9.29	0.80	0.07	2.62	0.74
	0.05	4.95	4.60	0.61	0.09	1.57	0.65
	0.10	2.24	3.15	0.54	0.09	1.18	0.47

It is not clear how to choose an appropriate level of rounding. This demonstrates the necessity for an efficient estimation method that takes measurement error into account.

4.2 Moving-resting Process with Measurement Error

Suppose the observations are recorded at times $t_0 = 0, t_1, \ldots, t_n$. Let $\{\epsilon_k\}_{k=0,\ldots,n}$ be independent and identically normally distributed random variables with mean 0 and variance σ_ϵ^2). An MR process with measurement error (MRME) $Z(t)$ at each time point is the superimposition of a measurement error and the exact location. That is, the

observed location at t_k, $Z(t_k)$, is

$$Z(t_k) = X(t_k) + \epsilon_k. \tag{4.1}$$

The coordinates of realizations from a two-dimensional moving-resting process with Gaussian measurement errors are plotted in the bottom right panel ($\sigma_\epsilon = 0.01$) and the bottom left panel ($\sigma_\epsilon = 0.05$) in Figure 5. As we mentioned in the introduction, the process $Z(t_k)\}_{k=0,\ldots,n}$ is not Markov. Moreover, even the location-state process $\{Z(t_k), S(t_k)\}_{k=0,\ldots,n}$ is not Markov. However, the discrete time process, $\{Z(t_{2k}) - Z(t_{2k-1}), S(t_{2k})\}_{k=1,\ldots,[n/2]}$, is Markov. More specifically, we will show below that the distribution of $(Z(t_{2k+2}) - Z(t_{2k+1}), S(t_{2k+2}))$ depends only on state $S(t_{2k})$.

First, let us calculate the marginal distribution of $Z(t) - Z(0)$ (the increment of $\{Z(t)\}_{t\geq 0}$ from time 0 to time t). Consider $\Delta Z(t) = Z(t) - Z(0) = X(t) - X(0) + \xi$, where $\xi \sim N(0, 2\sigma_\epsilon^2)$ independent of process $(X(t), S(t))$. Note that, $Z(t_{2k+2}) - Z(t_{2k+1}) = X(t_{2k+2}) - X(t_{2k+1}) + \epsilon_{2k+2} - \epsilon_{2k+1}$, and $\{\epsilon_{2k+2} - \epsilon_{2k+1}\}\sim N(0, 2\sigma_\epsilon^2)$. Without loss of generality, we assume that $X(0) = 0$. Denote

$$g_{ij}(z, t) = \mathbf{P}_i(\Delta Z(t) \in dz, S(t) = j)/\mathrm{d}z,$$

where $i, j = \{1, 0\}$. Then, we get that

$$g_{11}(z, t)\mathrm{d}z = \mathbf{P}_1(\Delta Z(t) \in \mathrm{d}z, S(t) = 1)$$

$$
\begin{aligned}
&= \int_{\mathbb{R}} \mathbf{P}_1(X(t) + \xi \in \mathrm{d}z, S(t) = 1, \xi \in \mathrm{d}x) \\
&= \int_{\mathbb{R}} \mathbf{P}_1(X(t) \in \mathrm{d}z - x, S(t) = 1)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x \\
&= \int_{\mathbb{R}} h_{11}^*(z - x, t)\mathrm{d}z\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x,
\end{aligned}
$$

where, as before, $\phi(\cdot; \sigma^2)$ is the density function of $N(0, \sigma^2)$. Similarly, one can get that

$$
\begin{aligned}
g_{10}(z, t) &= \int_{\mathbb{R}} h_{10}^*(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x, \\
g_{01}(z, t) &= \int_{\mathbb{R}} h_{01}^*(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x, \\
g_{00}(z, t) &= \int_{\mathbb{R}} h_{00}^*(z - x, t)\phi(x; 2\sigma_\epsilon^2)\mathrm{d}x + e^{-\lambda_0 t}\phi(z; 2\sigma_\epsilon^2).
\end{aligned}
$$

Next, let us denote

$$
\tau_{ij}(t) = \mathbf{P}_i(S(t) = j).
$$

It is easy to see

$$
\begin{aligned}
\tau_{01}(t) &= \mathbf{P}_0(S(t) = 1) \\
&= \sum_{n=0}^{\infty} \left[\mathbf{P}\left(\sum_{k=1}^{n+1} R_k + \sum_{k=1}^{n} M_k \le t \right) - \mathbf{P}\left(\sum_{k=1}^{n+1} R_k + \sum_{k=1}^{n+1} M_k \le t \right) \right] \\
&= \sum_{n=0}^{\infty} H(t, n, \lambda_1, n+1, \lambda_0).
\end{aligned}
$$

Here, $\{M_i\}_{i\ge1}$ and $\{R_i\}_{i\ge1}$ are defined in Section 4.1, and a summation over an empty set is 0. The functions $H(t, n, \lambda_1, n+1, \lambda_0)$ has been defined in Chapter 3.2 and were

investigated in Appendix A, where a computationally efficient algorithm for computing these functions was derived. Using similar technique, we obtain that

$$\tau_{10}(t) = \sum_{n=0}^{\infty} H(t, n, \lambda_0, n+1, \lambda_1),$$
$$\tau_{00}(t) = \sum_{n=0}^{\infty} H(t, n, \lambda_0, n, \lambda_1),$$
$$\tau_{11}(t) = \sum_{n=0}^{\infty} H(t, n, \lambda_1, n, \lambda_0).$$

Finally, we are ready to present the transition density from $S(0)$ to $(Z(t)-Z(u), S(t))$, where $0 < u < t$, $Z(u) = X(u) + \xi$, $Z(t) = X(t) + \eta$, and $\xi, \eta \overset{i.i.d.}{\sim} N(0, \sigma_\epsilon^2)$ that are also independent of process $(X(t), S(t))$. Using Markov property of the location-state process $(X(t)), S(t))$ and independence of the added noise, one can get that

$$f(Z(t) - Z(u), S(t) = j | S(0) = i) = \sum_{k=0}^{1} \tau_{ik}(u) g_{kj}(Z(t) - Z(u), t - u),$$

where $i, j = \{0, 1\}$.

In conclusion, let us note that real-world data sets are two-dimensional. The formulas from above can be generalized to a d-dimensional case. Let $\mathbf{X}(t_k)$ and $\mathbf{Z}(t_k)$ be d-dimensional random vectors. Set $\mathbf{X}(t_k) = (X_1(t_k), \ldots, X_d(t_k))$, $\epsilon_\mathbf{k} = (\epsilon_{1k}, \ldots, \epsilon_{dk}) \sim MN(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I})$, and $\epsilon_\mathbf{i}$ and $\epsilon_\mathbf{j}$ are independent for $i \neq j$. Then, $\mathbf{Z}(t_k) = (Z_1(t_k), \ldots, Z_d(t_k)) =$

$(X_1(t_k)+\epsilon_{1k},\ldots,X_d(t_k)+\epsilon_{dk})$. The density h_{ij} in the d-dimension case is given by

$$h^*_{10}(\mathbf{x},t)=\int_0^t\prod_{i=1}^d\phi(x_i;\sigma^2w)p^*_{10}(w,t)\mathrm{d}w.$$

Then we get that

$$\begin{aligned}
g_{10}(\mathbf{z},t)&=\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}h^*_{10}(\mathbf{z}-\mathbf{x};\sigma^2w)\prod_{i=1}^d\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_1\ldots\mathrm{d}x_d\\
&=\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\left[\int_0^t\prod_{i=1}^d\phi(z_i-x_i;\sigma^2w)p^*_{10}(w,t)\mathrm{d}w\right]\prod_{i=1}^d\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_1\ldots\mathrm{d}x_d\\
&=\int_0^t\prod_{i=1}^d\left[\int_{\mathbb{R}}\phi(z_i-x_i;\sigma^2w)\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_i\right]p^*_{10}(w,t)\mathrm{d}w.
\end{aligned}$$

Similarly, we also have

$$\begin{aligned}
g_{00}(\mathbf{z},t)&=\int_0^t\prod_{i=1}^d\left[\int_{\mathbb{R}}\phi(z_i-x_i;\sigma^2(t-w))\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_i\right]p^*_{00}(w,t)\mathrm{d}w+e^{-\lambda_0t}\prod_{i=1}^d\phi(z_i,2\sigma^2_\epsilon),\\
g_{01}(\mathbf{z},t)&=\int_0^t\prod_{i=1}^d\left[\int_{\mathbb{R}}\phi(z_i-x_i;\sigma^2(t-w))\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_i\right]p^*_{01}(w,t)\mathrm{d}w,\\
g_{11}(\mathbf{z},t)&=\int_0^t\prod_{i=1}^d\left[\int_{\mathbb{R}}\phi(z_i-x_i;\sigma^2w)\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_i\right]p^*_{11}(w,t)\mathrm{d}w\\
&\quad+e^{-\lambda_1t}\prod_{i=1}^d\left[\int_{\mathbb{R}}\phi(z_i-x_i;\sigma^2t)\phi(x_i;2\sigma^2_\epsilon)\mathrm{d}x_i\right].
\end{aligned}$$

Let us mention that these formulas do not require numerical multiple integral evaluation and, as a consequence, are computationally efficient.

4.3 Composite Likelihood Estimation

Since the full likelihood is unavailable, we resort to composite likelihood to estimate the parameters (Lindsay, 1988). A composite likelihood is a weighted product of likelihood segments

$$\text{CL} = \prod_{k=1}^{K} L_k^{w_k},$$

where L_k is the true likelihood of the k-th data segment with a non-negative weight w_k, $k = 1, \ldots, K$, and K is the number of segments depending on the construction of the CL. The weights can useful, for example, in pairwise likelihood when some pairs with stronger dependence contribute more than other pairs. Suppose that the location-state observations are denoted as

$$\mathbf{Z} = (Z(t_0), Z(t_1), \ldots, Z(t_n))$$

$$\mathbf{S} = (S(t_0), S(t_1), \ldots, S(t_n)).$$

The observed data only contains $\mathbf{Z}$. We propose two ways to construct composite likelihood.

4.3.1 Two-piece Composite Likelihood

The likelihood of increment-state observations at even time points

$$\mathbf{Z}_{even} = \left(Z_2, \ldots, Z_{2[n/2]}\right),$$

$$\mathbf{S}_{even} = \left(S(t_0), S(t_2), \ldots, S(t_{2[n/2]})\right),$$

where $Z_k = Z(t_k) - Z(t_{k-1})$ is given by

$$L\left(\mathbf{Z}_{even}, \mathbf{S}_{even}; \boldsymbol{\theta}\right) = \nu^*(S(t_0)) \prod_{k=1}^{[n/2]} f\left(Z_{2k}, S(t_{2k}) | S(t_{2k-2})\right),$$

where $[\cdot]$ is the integer function, $\boldsymbol{\theta} = (\lambda_1, \lambda_0, \sigma, \sigma_\epsilon)$ and $\nu^*(S(t_0))$ is the initial distribution that is assumed to be stationary. In practice, $\mathbf{S}_{even}$ are hidden states. So, we need the likelihood of the increment process $\mathbf{Z}_{even}$. This can obtained by taking sum over all possible state trajectories:

$$L(\mathbf{Z}_{even}; \boldsymbol{\theta}) = \sum_{S(t_0), S(t_2), \ldots, S(t_{2[n/2]}) \in \{0,1\}} L\left(\mathbf{Z}_{even}, \mathbf{S}_{even}; \boldsymbol{\theta}\right).$$

The cardinality of the set of the state trajectories is $2^{[n/2]+1}$. But it is still can be efficiently evaluated with help of dynamic programming, specifically, by the forward algorithm.

First, let us define the forward variables by

$$\alpha^*(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}) = \sum_{S(t_0),S(t_2),\ldots,S(t_{2k-2})\in\{0,1\}} \nu^*(S(t_0)) \times \prod_{j=1}^{k} f\left(Z(t_{2j}) - Z(t_{2j-1}), S(t_{2j})|S(t_{2j-2})\right),$$

where $\mathbf{Z}_{even}(t_{2k}) = (Z(t_0), Z(t_2), \ldots, Z(t_{2k}))$ and $k = 1, \ldots, [n/2]$, and the initial forward variable $\alpha^*\left(\mathbf{Z}_{even}(t_0), S(t_0), \boldsymbol{\theta}\right) = \nu^*(S(t_0))$. Then, it is easy to see that the forward variables satisfy the following recursive relationship:

$$\alpha^*\left(\mathbf{Z}_{even}(t_{2k+2}), S(t_{2k+2}), \boldsymbol{\theta}\right) = \sum_{S(t_{2k})\in\{0,1\}} \alpha^*\left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) \times f\left(\mathbf{Z}_{even}(t_{2k+2}) - \mathbf{Z}_{even}(t_{2k+1}), S(t_{2k+2})|S(t_{2k})\right).$$

This allows us to compute the likelihood in linear time with respect to n time, because

$$L(\mathbf{Z}_{even}; \boldsymbol{\theta}) = \sum_{S(t_{2[n/2]})\in\{0,1\}} \alpha^*\left(\mathbf{Z}_{even}(t_{2[n/2]}), S(t_{2[n/2]}), \boldsymbol{\theta}\right).$$

Now, when the sample size n is large, the likelihood tends to be too small to be distinguished from zero by computer. To address the underflow problem, one should use the normalized forward algorithm. More specifically, let us introduce the normalized

forward variables as

$$\bar{\alpha}^* \left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) = \frac{\alpha^* \left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right)}{L\left(\mathbf{Z}_{even}(t_{2k}); \boldsymbol{\theta}\right)},$$

and let

$$d^* \left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right) = \frac{L\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right)}{L\left(\mathbf{Z}_{even}(t_{2k}); \boldsymbol{\theta}\right)}.$$

Then, the update formulas for normalized forward variable $\bar{\alpha^*} \left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right)$ and $d\left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right)$ are given by

$$\begin{aligned}\bar{\alpha}^* \left(\mathbf{Z}_{even}(t_{2k+2}), S(t_{2k+2}), \boldsymbol{\theta}\right) =& \frac{1}{d^* \left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right)} \\ & \times \sum_{S(t_{2k}) \in \{0,1\}} \bar{\alpha}^* \left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) \\ & \times f\left(\mathbf{Z}_{even}(t_{2k+2}) - \mathbf{Z}_{even}(t_{2k+1}), S(t_{2k+2}) | S(t_{2k})\right),\end{aligned}$$

and

$$\begin{aligned}d^* \left(\mathbf{Z}_{even}(t_{2k+2}); \boldsymbol{\theta}\right) = & \sum_{S(t_{2k+2}), S(t_{2k}) \in \{0,1\}} \bar{\alpha}^* \left(\mathbf{Z}_{even}(t_{2k}), S(t_{2k}), \boldsymbol{\theta}\right) \\ & \times f\left(\mathbf{Z}_{even}(t_{2k+2}) - \mathbf{Z}_{even}(t_{2k+1}), S(t_{2k+2}) | S(t_{2k})\right).\end{aligned}$$

Finally, the likelihood function is given by

$$\log L\left(\mathbf{Z}_{even}(t_{2k});\boldsymbol{\theta}\right)=\sum_{k=1}^{[n/2]}\log d^{*}\left(\mathbf{Z}_{even}(t_{2k});\boldsymbol{\theta}\right). \tag{4.2}$$

In a similar fashion, one can compute the likelihood of the observed increments at the odd time points $\mathbf{Z}_{odd}=\left(Z_1,Z_3,\ldots,Z_{2[(n+1)/2]-1}\right)$. Adding two log-likelihoods together we get the following composite log-likelihood:

$$cl\left((Z(t_0),\ldots,Z(t_n));\boldsymbol{\theta}\right)=\log\left(L(\mathbf{Z}_{even};\boldsymbol{\theta})\right)+\log\left(L(\mathbf{Z}_{odd};\boldsymbol{\theta})\right). \tag{4.3}$$

The maximum composite likelihood estimator (MCLE) of $\boldsymbol{\theta}$ is the maximizer $\hat{\boldsymbol{\theta}}$ of (4.3).

4.3.2 Marginal Composite Likelihood

The second approach is to use the one-step transition density with the dependence between two consecutive increments discarded. If $\mathbf{S}$ were observed, for $i,j\in\{0,1\}$, the likelihood of each pair of consecutive location-state observations

$$(\{Z(t_{k-1}),S(t_{k-1})=i\},\{Z(t_k),S(t_k)=j\})$$

is

$$\nu^{*}(S(t_{k-1})=i)g_{ij}(Z(t_k)-Z(t_{k-1}),t_k-t_{k-1}),$$

where $\nu^*(\cdot)$ is the stationary distribution of state process $\{S(t)\}_{t\geq 0}$. Since $\mathbf{S}$ is unobserved, the likelihood of $(Z(t_{k-1}), Z(t_k))$ is

$$\sum_{j=0}^{1}\sum_{i=0}^{1}\nu^*(S(t_{k-1})=i)g_{ij}(Z(t_k)-Z(t_{k-1}),t_k-t_{k-1}).$$

The marginal composite log-likelihood is

$$\begin{aligned} & cl^*((Z(t_0),\ldots,Z(t_n));\boldsymbol{\theta}) \\ =& \sum_{k=1}^{n}\log\left[\sum_{j=0}^{1}\sum_{i=0}^{1}\nu^*(S(t_{k-1})=i)g_{ij}(Z(t_k)-Z(t_{k-1}),t_k-t_{k-1})\right]. \end{aligned} \tag{4.4}$$

Since the dependence among the increments is descarded, the resulting estimator is expected to be less efficient if the dependence is stronger.

4.3.3 Variance Estimation

To make inferences about $\boldsymbol{\theta}$, we need the variance of $\hat{\boldsymbol{\theta}}$. It can be estimated by parametric bootstrap with the time points fixed easily because simulating from the MRME process is simple. The general approach of parametric bootstrap is given as Algorithm 1.

Alternatively, we can estimate the variance by inverting the observed Godambe information matrix (Godambe, 1960)

$$G(\boldsymbol{\theta}) = H(\boldsymbol{\theta})J(\boldsymbol{\theta})^{-1}H(\boldsymbol{\theta}),$$

Algorithm 1: Estimated standard error from parametric bootstrap

input : Observed data; number of resampling M.
· Fit model to get the parameter estimates;
for *m = 1 to M* **do**
  · Use the estimated parameters to generate a bootstrap sample on the observed time grids;
  · Fit model to the bootstrap sample to get bootstrap estimate;
end
· Make inference based on empirical distribution from these bootstrap estimators.

where

$$H(\boldsymbol{\theta}) = \mathbf{E}\left[-\frac{\partial^2}{\partial \boldsymbol{\theta}^2} cl((Z(t_0), \ldots, Z(t_n)); \boldsymbol{\theta})\right],$$

and

$$J(\boldsymbol{\theta}) = \mathrm{Var}\left[\frac{\partial}{\partial \boldsymbol{\theta}} cl((Z(t_0), \ldots, Z(t_n)); \boldsymbol{\theta})\right].$$

Practically, $H(\boldsymbol{\theta})$ is estimated by the Hessian matrix of the negative composite likelihood evaluated at $\hat{\boldsymbol{\theta}}$. Calculation of $J(\boldsymbol{\theta})$ is more difficult as there is no replicated data to estimate this variance. Parametric bootstrap can be applied to evaluate $J(\boldsymbol{\theta})$ as the empirical variance of gradient of composite likelihood from a large number of bootstrap samples. Finally, $\mathrm{Var}(\hat{\boldsymbol{\theta}})$ can be obtained by the inverse of $G(\hat{\boldsymbol{\theta}})$ (e.g., Varin et al., 2011). However, according to our simulation result, this approach does not perform as good as parametric bootstrap approach.

4.4 Simulation Study

We ran several simulations to check performance of the MCLE based on both the marginal composite likelihood and the two-piece composite likelihood. The objective of this study is threefold: (1) to see if the procedures successfully recovers the model parameters, (2) to verify that standard errors can be obtained with help of parametric bootstrap, and (3) to compare performance of the marginal method to the two-piece one.

First, we generated movement data using MRME model described by equation (4.1). The model parameters were set to be $\lambda_1 = 1$, $\lambda_0 = 0.5$, $\sigma = 1$, and $\sigma_\epsilon \in (0.01, 0.05)$. This is the same setup that was used for simulations in Section 4.1. For each configuration, we generated 200 two-dimensional datasets on a time grid from 0 to 1000, with sampling interval 5. The resulting data has length $n = 200$.

Figure 6 presents the violin plots of the composite likelihood estimates of the 200 replicates in comparison to the true values of the four parameters. Violin plots are similar to box plots with a rotated kernel density plot on each side. The horizontal bars in the panels are the true parameter values. For each parameter, the true value lies in the bulk part of the violin plot. This indicates that the true parameters are recovered well by both MCLE methods. The left and right panels have different size of added measurement errors. The variation of the estimates in the case of $\sigma_\epsilon = 0.01$ is noticeably smaller than that in the case of $\sigma_\epsilon = 0.05$, which is expected.

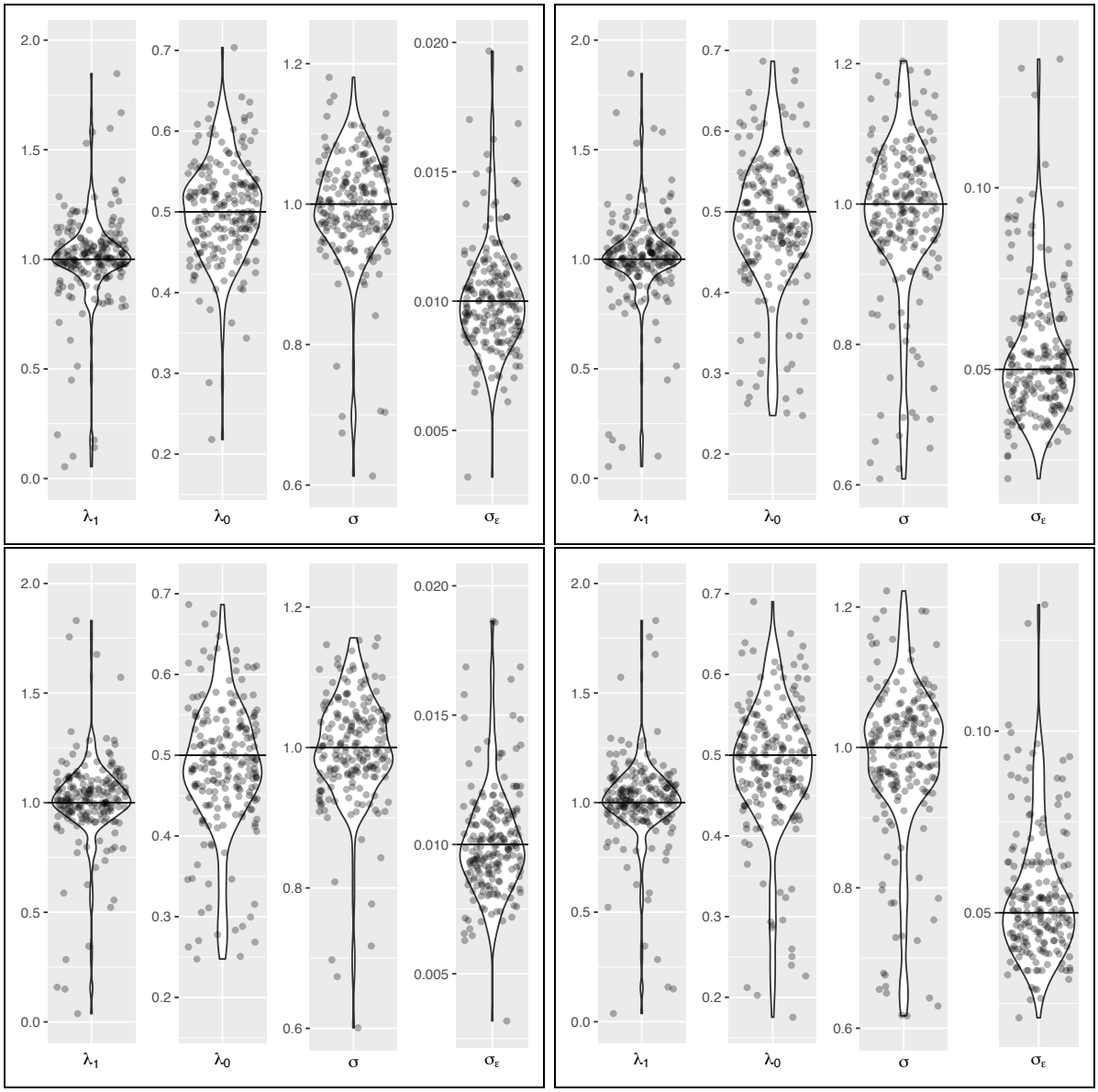

Figure 6: Violin plots of the MCLE with two-piece method (*top*) and marginal method (*bottom*) from 200 replicates. The horizontal bar in each panel is the true parameter value $\lambda_1 = 1$, $\lambda_0 = 0.5$, $\sigma = 1$, and $\sigma_\epsilon = 0.01$ (*left*) and 0.05 (*right*). The number of replications is 200.

The second simulation study addresses the problem of estimating of standard errors for both MCLE procedures via parametric bootstrap. The sampling horizon (the length of observation window) is set to two levels, 200 and 500 time units. The sampling interval (the inverse sampling frequency) also have two levels, 1 and 5 time units. The parameters of MRME process are: $\lambda_1 = 1$, $\lambda_0 = 0.5$, $\sigma = 1$, and $\sigma_\epsilon = 0.01$. The result of simulation is recorded in Table 5. Once again we can see that both the marginal method and the two-piece method recover the true parameters well. Their empirical standard errors are similar. This tells us that two methods have comparable efficiency for these setups. Moreover, the standard errors were estimated by parametric bootstrap procedure. As we can see, the estimated standard errors are reasonably close to empirical ones.

Finally, let us note that the performance of two methods is similar in two simulation studies described above. It is a bit surprising because the marginal method basically ignores the dependence of the MRME process and treats its increment as they are independent. One possible explanation is that if the distance between two consecutive observation is relatively long then the dependence between two consecutive increments of the MRME process is weaker. To see that one can easily calculate the correlation of absolute values of consecutive increments via simulation by employing the auto-correlation function with lag 1 (ACF(1)). For example, for the same parameter set as in the above simulations and a long time horizon 100000, the ACF(1) for the sampling interval 5 is 0.02, but the ACF(1) for the sampling interval 0.1 is 0.46. The results of simulation

Table 5: Summaries of average estimator (EST), empirical standard error (ESE), and average parametric bootstrap standard error (ASE) of maximum composite likelihood estimator with two-piece method (Formula (4.3)) and marginal mrthod (Formula (4.4)). The number of replications is 200.

Sampling	Sampling	Parameter	True	Two-piece method			Marginal method		
horizon	interval		value	EST	ESE	ASE	EST	ESE	ASE
200	1	λ_1	1.0	1.104	0.394	0.386	1.057	0.410	0.414
		λ_0	0.5	0.502	0.093	0.092	0.488	0.109	0.109
		σ	1.0	1.011	0.084	0.088	1.001	0.089	0.097
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.002	0.070	0.068	1.002	0.069	0.068
	5	λ_1	1.0	0.961	0.546	0.508	0.982	0.570	0.516
		λ_0	0.5	0.493	0.169	0.240	0.485	0.211	0.251
		σ	1.0	0.966	0.189	0.163	0.970	0.194	0.166
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.145	0.710	0.651	1.136	0.665	0.648
500	1	λ_1	1.0	1.038	0.240	0.224	0.983	0.225	0.223
		λ_0	0.5	0.508	0.060	0.058	0.495	0.065	0.062
		σ	1.0	1.008	0.060	0.055	0.997	0.067	0.058
		$\sigma_\epsilon(\times 10^{-2})$	1.0	0.998	0.044	0.042	0.998	0.044	0.042
	5	λ_1	1.0	1.020	0.362	0.342	1.009	0.354	0.335
		λ_0	0.5	0.512	0.101	0.111	0.509	0.106	0.115
		σ	1.0	0.982	0.122	0.114	0.978	0.119	0.112
		$\sigma_\epsilon(\times 10^{-2})$	1.0	1.046	0.356	0.376	1.045	0.362	0.357

Table 6: Summaries of average estimator (EST) and empirical standard error (ESE) of maximum composite likelihood estimator for two-piece method (Formula (4.3)) and marginal method (Formula (4.4)). The number of replications is 200.

Sampling horizon	Sampling interval	Parameter	True value	Two-piece method		Marginal method	
				EST	ESE	EST	ESE
200	0.1	λ_1	1.0	1.031	0.238	1.132	0.329
		λ_0	0.1	0.102	0.024	0.109	0.026
		σ	1.0	1.002	0.040	1.007	0.043
		$\sigma_\epsilon(\times 10^{-2})$	1.0	0.999	0.015	0.999	0.015

with small sampling interval 0.1 are presented in Table 6. This simulation suggests that the two-piece procedure is preferable for datasets with shorter sampling intervals (more frequent observations).

4.5 Movement of Mountain Lion

The MRME model was applied to the global positioning system (GPS) data of a mature female mountain lion living in the Gros Ventre Mountain Range near Jackson Wyoming. The data were collected by a code-only GPS wildlife tracking collar from 2009 to 2012. The collar was designed to record the location every 8 hours, but the actual sampling intervals were irregular. The movement behaviors of mountain lions are known to be different across seasons. So, we fitted a MRME model to the summer data (from June 1, 2012 to August 31, 2012) and winter data (from December 1, 2011 to February 29, 2012) separately. These two periods of data were plotted as Figure 7. The summer data had an average sampling interval of 5.46 hours with standard deviation 5.14 hours,

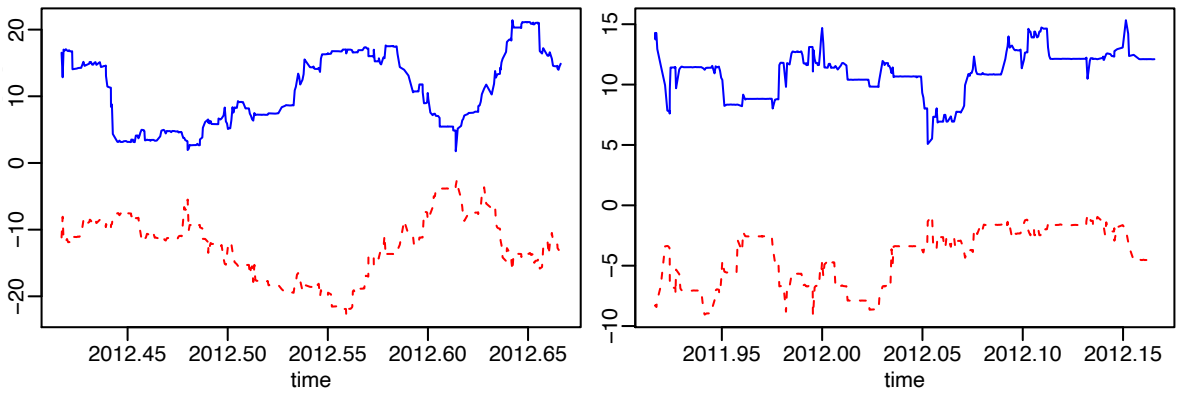

Figure 7: Actual coordinates of a female mountain lion in the Gros Ventre mountain range, Wyoming. The x-axis is time in years. The y-axis is departure from the starting point. The solid blue line is UTM easting (km) and the dashed red line is UTM northing (km). *Left:* Summer period data, from June 1, 2012 to August 31, 2012. *Right:* Winter period data, from December 1, 2011 to February 29, 2012.

ranging from 0.5 hours to 32 hours. The average sampling interval is 5.58 hours in the winter data, with standard deviation 4.09 hours and range from 0.5 hours to 25 hours. The summer data has 401 observations and winter data has 392 observations.

The maximum two-piece composite likelihood (Formula (4.3)) estimates for summer data are $\hat{\lambda_1} = 2.841/hour$, $\hat{\lambda_0} = 0.179/hour$, $\hat{\sigma} = 1.335km/hour^{1/2}$ and $\hat{\sigma_\epsilon} = 0.019km$. On average, this female mountain lion stays in moving and resting for 0.352 hours and 5.587 hours during summer and if it keeps moving for one hour, the average deviation from the initial position is $1.335km$ in both directions (northing and easting). Compared to summer data, the estimates for winter data are $\hat{\lambda_1} = 6.225/hour$, $\hat{\lambda_0} = 0.118/hour$, $\hat{\sigma} = 1.506km/hour^{1/2}$ and $\hat{\sigma_\epsilon} = 0.009km$. It is clearly to see that during winter period this female mountain lion spends 51.7% more time in each staying and 54.4% less time in each moving. The mobility keeps comparable. The estimate of σ_ϵ (standard deviation

Table 7: Analysis results for mountain lion movement data. Point estimates (EST) from both two-piece method and marginal method are reported. Standard error of point estimates are evaluated by parametric bootstrap (SE).

Season	Parameter	Two-piece method		Marginal method	
		EST	SE	EST	SE
Summer	λ_1	2.841	0.459	1.090	0.280
	λ_0	0.179	0.014	0.158	0.015
	σ	1.335	0.106	0.999	0.104
	$\sigma_\epsilon(\times 10^{-2})$	1.854	0.087	1.879	0.078
Winter	λ_1	6.225	0.825	4.720	0.711
	λ_0	0.118	0.010	0.114	0.009
	σ	1.506	0.095	1.454	0.089
	$\sigma_\epsilon(\times 10^{-2})$	0.908	0.036	0.934	0.043

of Gaussian noise) indicates that the GPS tracking collar has about 10 to 20 meters measurement error and the error is twice higher in summer than it in winter, that is consistent to the knowledge from ecologist (Owari et al., 2009). The result from marginal method is similar (Table 3).

Chapter 5

Concluding Remarks

For moving-resting-handling process, the methodology developed in Sections 3.2 works even if the holding times are not exponentially distributed, which is an advantage of our approach. If we want to keep the Markov property, then all holding times must have exponential distributions. The memoryless distribution might be not appropriate for some species that follow a cyclic daily routine. Nonetheless, if animals under observation do not exhibit a daily periodic behavior (like mountain lions), then using an exponential distribution is acceptable. The behavior of these animals is subject to interruptions that can cut their time spent in a particular activity. For example, handling might be interrupted by a more dominate predator who drives the lion off her kill before she is finished with it.

A different (from exponential) distribution should be used for species with a periodic routine. One interesting possibility is to employ stable distributions (for example, Lévy distribution). Because a linear combination of two independent random variables with a stable distribution has the same distribution, up to location and scale parameters, the formulas in Theorem 3 will be even nicer. The drawback is that the state process

is then semi-Markov, and, as a result, the likelihood inferences from standard HMM tools are not available. Nevertheless, this still might be of interest for practitioners, because estimation can be done via alternative methods such as the composite likelihood estimation (Lindsay, 1988).

Considering moving-resting process with measurement error, inactive periods and measurement errors are both desired features of animal movement models. Handling measurement errors added to the MR process is especially critical because, if discarded, a microscopic amount of measurement error would cause substantial bias in estimation. Rounding the observed data is too ad hoc. The issue is rooted in that the flat pieces in the trajectory of an MR process would disappear when noises are present. Our approach based composite likelihood is the first to make the MRME model practically feasible. For movement data from predators that are known to have long inactive periods, the MRME model has great potential in revealing insights for animal ecologists.

The MRME model can be extended to meet further practical needs. Introducing measurement errors to the MRH model would be useful, since an inactive period may have different purposes for some predators. Nonetheless, evaluating the likelihood of the MRH model is extremely computationally demanding. The moving period may be of different types too to accommodate different moving behaviors (Benhamou, 2011; Kranstauber et al., 2012). This can be done by allowing the volatility parameter to have multiple levels. More generally, it can depend on terrain, weather conditions, or other covariates. Our MRME model provides a benchmark for these more realistic extensions.

Appendix A

Density and Distribution Evaluation for Convolution of Independent Gamma Variables

A.1 Introduction

Let $X_1, \ldots, X_n$ be n mutually independent random variables that have gamma distributions with shape parameters $\alpha_i > 0$ and scale parameters $\beta_i > 0$, $j = 1, \ldots, n$. Then random variable $Y = \sum_{i=1}^{n} X_i$ is the convolution of independent gamma variables. Another common parametrization of Y is weighted gamma convolution, that is $Y = \sum_{i=1}^{n} \beta_i Z_i$, where β_i are weights and Z_i are mutually independent gamma random variable with different shape parameters α_i and the same scale parameters as 1. The advantage of this expression is better and more explicit practical interpretation. For example, the compound weighted gamma convolution (Di Salvo, 2008), can be established by assuming random weights β_i. Without loss of generality, the scale parameters

β_i's can be assumed to be distinct. If it is not a case, variables with the same scales can be summed before the convolution. The stochastic properties of the convolution have been extensively studied in the literature (Khaledi and Kochar, 2013). Evaluation of the density and distribution of Y, however, has not been available in any standard statistical software package in spite of their importance in the aforementioned real world applications.

Exact evaluations of the density and distribution function of Y have no simple closed-forms and are challenging. Mathai (1982) was the first to express the density in terms of confluent hypergeometric functions of $n-1$ variables with arbitrary shape and scale parameters. When $n = 2$, the confluent hypergeometric function is univariate, and an efficient implementation is available in the GNU Scientific Library (GSL) (Galassi et al., 2009). Mathai (1982) also gave relatively easier-to-compute expressions in the special cases when all shapes are integers or are identical. Moschopoulos (1985) simplified the complex expression into a single gamma-series representation and provided a formula for the truncation error to evaluate the precision of numerical computation. Akkouchi (2005) indicated that the density function can be expressed in terms of an integral of the generalized beta function, but did not give explicit ways to numerically evaluate the integral. Di Salvo (2008) characterized the density function of Y as the product of a gamma density and a confluent form of the fourth Lauricella function (Srivastava and Karlsson, 1985), and introduced an accurate procedure to compute the fourth Lauricella function. The moments, cumulative distribution function, and related properties of Y were also

discussed in Di Salvo (2008). Vellaisamy and Upadhye (2009) proposed a random parameter representation of the gamma-series of Moschopoulops' method (Moschopoulos, 1985), where the weights define the probability mass function of a discrete distribution on non-negative integers. Implementations of the gamma-series methods are relatively simple but the computation is demanding when the variability of the scale parameters is large and the shape parameters are small. Built on the representation of Vellaisamy and Upadhye (2009), Barnabani (2017) proposed a fast approximation which approximates the weights of the gamma-series by a discrete distribution.

The contribution of this article is two-fold. First, we give a review of the exact methods of Mathai (1982) and Moschopoulos (1985). These methods are implemented in our R package `coga` (Hu et al., 2019). Their computational efficiency are compared in cases of $n = 2$ and $n = 3$. In the case of $n = 3$, the accuracy of the fast approximation from Barnabani (2017) is also assessed using our implementation. Our second contribution is an application to a renewal process with holding times following a mixture of exponential distributions. A new formula based on the confluent hypergeometric function for the probability mass function of the number of renewals by a given amount of time is derived. The computational efficiency of the new formula is shown via a numerical study.

A.2 Exact Evaluations

Let us introduce our notation first. Let $G(y;\alpha,\beta)$ and $g(y;\alpha,\beta)$ be, respectively, the distribution function and density function of a gamma variable with shape parameter α and scale parameter β. For the special case where $\beta = 1$, we use $G(y;\alpha)$ and $g(y;\alpha)$, respectively. Let $F(x;(\alpha_1,\beta_1),(\alpha_2,\beta_2))$ and $f(x;(\alpha_1,\beta_1),(\alpha_2,\beta_2))$, respectively, be the distribution and density function of Y in the case of $n = 2$. Finally, let $(x)_m = x(x+1)\ldots(x+m-1)$, which is the Pochhammer polynomial (Abramowitz and Stegun, 1972, Equation 6.1.22).

A.2.1 Mathai's Method

Mathai (1982) expresses the density of Y as

$$\begin{aligned} f(x) = &\left[\prod_{j=1}^{n} \beta_j^{\alpha_j}\Gamma(\gamma)\right]^{-1} x^{\gamma-1}e^{-x/\beta_1} \\ &\times \phi\Big(\alpha_2,\ldots,\alpha_n;\gamma;(1/\beta_1-1/\beta_2)x,\ldots,(1/\beta_1-1/\beta_n)x\Big), \end{aligned} \tag{A.1}$$

where $\beta_1 = \min_j(\beta_j)$, $\gamma = \sum_{j=1}^{n}\alpha_j$, and ϕ is a confluent hypergeometric function of $n-1$ variables. The confluent hypergeometric function is a special function given by the following multiple series:

$$\begin{aligned} &\phi\Big(\alpha_2,\ldots,\alpha_n;\gamma;(1/\beta_1-1/\beta_2)x,\ldots,(1/\beta_1-1/\beta_n)x\Big) \\ =&\sum_{r_2=0}^{\infty}\cdots\sum_{r_n=0}^{\infty}\left\{\frac{(\alpha_2)_{r_2}\ldots(\alpha_n)_{r_n}}{r_2!\ldots r_n!(\gamma)_r}\left[(1/\beta_1-1/\beta_2)x\right]^{r_2}\ldots\left[(1/\beta_1-1/\beta_n)x\right]^{r_n}\right\}. \end{aligned} \tag{A.2}$$

This function has been extensively studied (Mathai and Saxena, 1978). An alternative derivation of Equation (A.1) is given by Di Salvo (2008). The key idea is to express the confluent hypergeometric function as an integral over the standard hypercube with the integrand that poses good analytical properties. For completeness, we illustrate the derivation given by Di Salvo (2008) here. Equation (A.2) can be expressed as a multiple integral over $n-1$-dimensional standard simplex (Srivastava and Karlsson, 1985), that is,

$$\begin{aligned}\phi = \frac{\Gamma(\gamma)}{\prod_{i=1}^{n}\Gamma(\alpha_i)} \int_{E_{n-1}} \exp\{x\sum_{i=2}^{n} t_i(\frac{1}{\beta_1} - \frac{1}{\beta_i})\} \\ \times \prod_{i=2}^{n} t_i^{\alpha_i - 1}(1 - t_2 - \cdots - t_n)^{\alpha_1 - 1}\mathrm{d}t_2 \ldots \mathrm{d}t_n,\end{aligned} \tag{A.3}$$

where $E_{n-1} = \{(t_2, \ldots, t_n) : t_2 > 0, \ldots, t_n > 0, t_2 + \cdots + t_n \leq 1\}$. Di Salvo (2008) demonstrated that, with transformation

$$\begin{aligned} t_2 &= u_2 \\ t_3 &= u_3(1 - u_2) \\ t_4 &= u_4(1 - u_2)(1 - u_3) \\ &\ldots \\ t_n &= u_n(1 - u_2)(1 - u_3)\ldots(1 - u_{n-1}), \end{aligned}$$

Equation (A.3) can be written as a multiple integral over the standard hypercube, that is,

$$\begin{aligned}\phi = \frac{\Gamma(\gamma)}{\prod_{i=1}^{n}\Gamma(\alpha_i)} \int_{[0,1]^{n-1}} & \exp\{x\sum_{i=2}^{n} u_i(\frac{1}{\beta_1} - \frac{1}{\beta_i})h_i(\mathbf{u})\} \\ & \times \prod_{i=2}^{n} u_i^{\alpha_i - 1}(1-u_i)^{\bar{\alpha_i}} \mathrm{d}u_2 \ldots \mathrm{d}u_n,\end{aligned} \tag{A.4}$$

where $h_2 = 1$, $h_i(\mathbf{u}) = \prod_{j=2}^{i-1}(1-u_j)$, $i = 3, \ldots, n$, and $\bar{\alpha_i} = \sum_{j=i+1}^{n}\alpha_j + \alpha_1 - 1$, $i = 2, \ldots, n-1$, $\bar{\alpha_n} = \alpha_1 - 1$. Because of good analytical properties of integrand in Equation (A.4), the multiple integral can be efficiently computed.

For the special case of $n = 2$, the density is expressed in terms of the Kummer's confluent hypergeometric function ${}_1F_1$ (Abramowitz and Stegun, 1972, Equation 13.1.2) as

$$\begin{aligned} f(x;(\alpha_1,\beta_1),(\alpha_2,\beta_2)) &= \frac{x^{\gamma-1}e^{-x/\beta_1}}{\beta_1^{\alpha_1}\beta_2^{\alpha_2}\Gamma(\gamma)} {}_1F_1(\alpha_2;\gamma;(1/\beta_1 - 1/\beta_2)x) \\ &= \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} g(x;\gamma,\beta_1) {}_1F_1(\alpha_2;\gamma;(1/\beta_1 - 1/\beta_2)x). \end{aligned} \tag{A.5}$$

The benefit of Equation (A.5) is that the GSL (Galassi et al., 2009) has an implementation of ${}_1F_1$. Note that the condition $\beta_1 < \beta_2$ is not needed in (A.5). Indeed, if $\beta_1 > \beta_2$, then

$$\begin{aligned} f(x;(\alpha_1,\beta_1),(\alpha_2,\beta_2)) &= f(x;(\alpha_2,\beta_2),(\alpha_1,\beta_1)) \\ &= \frac{x^{\gamma-1}e^{-x/\beta_2}}{\beta_1^{\alpha_1}\beta_2^{\alpha_2}\Gamma(\gamma)} {}_1F_1(\alpha_1;\gamma;(1/\beta_2 - 1/\beta_1)x) \\ &= \frac{x^{\gamma-1}e^{-x/\beta_2}}{\beta_1^{\alpha_1}\beta_2^{\alpha_2}\Gamma(\gamma)} e^{-(1/\beta_2 - 1/\beta_1)x} {}_1F_1(\alpha_2;\gamma;(1/\beta_1 - 1/\beta_2)x) \end{aligned}$$

$$= \frac{x^{\gamma-1}e^{-x/\beta_1}}{\beta_1^{\alpha_1}\beta_2^{\alpha_2}\Gamma(\gamma)} {}_1F_1(\alpha_2;\gamma;(1/\beta_1 - 1/\beta_2)x),$$

where the third equation follows from ${}_1F_1(a;b;z) \equiv e^z {}_1F_1(b-a;b;-z)$ (Abramowitz and Stegun, 1972, Equation 13.1.27).

Finally, the distribution function of Y can be expressed in terms of incomplete gamma functions (with the gamma function kernels $x^{\gamma+r_i-1}e^{-x/\beta_1}$) by term-by-term integration of Equation (A.1). In particular, the distribution function when $n = 2$ can be expressed as follows: for $y > 0$

$$\begin{aligned} & F(y;(\alpha_1,\beta_1),(\alpha_2,\beta_2)) \\ &= \frac{1}{\beta_1^{\alpha_1}\beta_2^{\alpha_2}} \sum_{k=0}^{\infty} \frac{\binom{\alpha_2+k-1}{k}}{\Gamma(\gamma+k)} (1/\beta_1 - 1/\beta_2)^k \int_0^y x^{k+\gamma-1}e^{-x/\beta_1}\mathrm{d}x \\ &= \frac{1}{\beta_1^{\alpha_1}\beta_2^{\alpha_2}} \sum_{k=0}^{\infty} \frac{\binom{\alpha_2+k-1}{k}}{\Gamma(\gamma+k)} (1/\beta_1 - 1/\beta_2)^k \beta_1^{k+\gamma} G(y/\beta_1;k+\gamma)\Gamma(k+\gamma) \\ &= \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} \sum_{k=0}^{\infty} \binom{\alpha_2+k-1}{k} (1-\beta_1/\beta_2)^k G(y/\beta_1;k+\gamma). \end{aligned} \tag{A.6}$$

The gamma distribution function G in the equation can be evaluated with standard statistical packages.

A.2.2 Moschopoulos' Method

Moschopoulos (1985) expresses the density of Y by a single gamma series with coefficients

that can be calculated recursively:

$$\begin{aligned} f(x) &= C \sum_{k=0}^{\infty} \delta_k x^{\rho+k-1} e^{-x/\beta_1} / \left[\Gamma(\rho+k) \beta_1^{\rho+k} \right] \\ &= C \sum_{k=0}^{\infty} \delta_k g(x; \rho+k, \beta_1), \qquad x > 0, \end{aligned}$$

where $\beta_1 = \min_i(\beta_i)$, $C = \prod_{i=1}^{n} (\beta_1/\beta_i)^{\alpha_i}$, $\rho = \sum_{i=1}^{n} \alpha_i > 0$, and δ_k are given by the recursive relations

$$\delta_{k+1} = \frac{1}{k+1} \sum_{i=1}^{k+1} i \gamma_i \delta_{k+1-i}, \qquad k = 0, 1, 2, \ldots,$$

with $\delta_0 = 1$ and $\gamma_k = \sum_{i=1}^{n} \alpha_i (1 - \beta_1/\beta_i)^k / k$ for $k = 1, 2, \ldots$. This expression facilitates distribution function evaluation as

$$F(y) = C \sum_{k=0}^{\infty} \delta_k G(y; \rho+k, \beta_1), \qquad y > 0.$$

The weights $C\delta_k$'s can be viewed as the probability masses of a discrete random variable on non-negative integers (Vellaisamy and Upadhye, 2009). When $n = 2$, this discrete distribution is negative binomial. For $n > 2$, Barnabani (2017) proposed to approximate the discrete distribution by a three-parameter generalized negative binomial distribution defined by Jain and Consul (1971) through moment matching.

Table 8: Timing comparison (in microseconds) of Mathai's and Moschopoulos' methods when $n = 2$ in evaluating the density and distribution function of convolutions of independent gamma variables.

Parameters			Density		Distribution	
α	β_1	β_2	Moschopoulos	Mathai	Moschopoulos	Mathai
0.2	0.4	0.3	23,030	86	25,011	2,648
	4	0.3	103,987	176	110,804	9,018
	4	3	24,566	88	25,696	2,849
2	0.4	0.3	29,163	94	31,030	3,086
	4	0.3	165,901	96	173,916	13,590
	4	3	30,378	90	33,489	3,397
20	0.4	0.3	53,231	103	57,468	6,471
	4	0.3	538,807	175	566,857	37,527
	4	3	53,259	108	58,479	6,604

A.2.3 Timing Comparison

We implemented the methods of Mathai (1982) and Moschopoulos (1985) in an open source R package `coga` (Hu et al., 2019). The computation is done in `C++` code and interfaced to R (R Core Team, 2018) in the `coga` package. In addition, the fast approximation from Barnabani (2017) is also available in the package for $n > 2$.

We first compare the speed of the two methods in the case of $n = 2$. Since the gamma-series of Moschopoulos (1985) is CPU-time intensive when the shape parameters are small and the scale parameters have large variation (Barnabani, 2017), we considered such situations and its complements. The shape parameters were set to be $\alpha_1 = \alpha_2 \in \{0.2, 2, 20\}$. The scale parameters were set to be $(\beta_1, \beta_2) \in \{(0.4, 0.3), (4, 0.3), (4, 3)\}$. For each configuration, we used a random sample of size $100,000$ from the distribution

to determine the bulk range of the observations. Then we evaluate the density and the distribution of the convolution over 100 equally spaced grid points in the bulk range. The evaluations were repeated 100 times.

Table 8 summarizes the median time to evaluate density and distribution at the 100-point grid from 100 replicates obtained on an Intel 2.50GHz computer. Density evaluation using Mathai's method (implemented with the ${}_1F_1$ function from the GSL) performs much faster than Moschopoulos' method in all settings; in some settings, it is up to 3,000 times faster. Distribution evaluation takes much longer than density evaluation using Mathai's method, but is still up to 16 times faster than Moschopoulos' method. Moschopoulos' method takes longer when the scale parameters are very different, $(\beta_1, \beta_2) = (4, 0.3)$, or the shape parameters bigger. Mathai's method is much less sensitive to the parameter settings.

Following the design and steps in the case of $n = 2$, we conducted a numerical analysis for $n = 3$. The shape parameters were set to be $\alpha_1 = \alpha_2 = \alpha_3 \in \{0.2, 2, 20\}$. The scale parameters were set to be

$$(\beta_1, \beta_2, \beta_3) \in \{(0.4, 0.3, 0.2), (4, 0.3, 0.2), (4, 3, 0.2), (4, 3, 2)\}.$$

The two methods gave numerically indistinguishable results in both density and distribution evaluations, verifying each other. Table 9 summarize the median timing results from 100 replicates. When $n > 2$, the multivariate confluent hypergeometric function

Table 9: Timing comparison (in milliseconds) of Mathai's, Moschopoulos' exact methods and Barnabani's approximation method when $n = 3$.

Parameters				Density			Distribution		
α	β_1	β_2	β_3	Moschopoulos	Mathai	Barnabani	Moschopoulos	Mathai	Barnabani
0.2	0.4	0.3	0.2	37	1,223	6	39	1,428	8
	4	0.3	0.2	167	5,796	18	181	8,067	28
	4	3	0.2	186	10,208	19	197	14,000	30
	4	3	2	38	1,245	6	40	1,474	8
2	0.4	0.3	0.2	48	1,044	8	51	1,328	11
	4	0.3	0.2	242	5,333	21	253	8,975	35
	4	3	0.2	313	12,597	25	331	20,646	43
	4	3	2	49	1,082	8	53	1,415	11
20	0.4	0.3	0.2	109	3,250	14	119	4,721	22
	4	0.3	0.2	780	16,083	29	950	40,704	78
	4	3	0.2	596	21,553	18	1,418	133,699	101
	4	3	2	110	3,329	14	123	4,953	23

has no efficient implementations yet to the best of our knowledge. Therefore, Mathai's method needs to evaluate $n - 1$ nested infinite series, which makes it quite complicated (Jasiulewicz and Kordecki, 2003; Sen and Balakrishnan, 1999). In this case, Moschopoulos' method is faster for its single-series. The approximation from Barnabani (2017) was also included in the comparison, which is up to 30 times faster in density evaluation and 14 times faster in distribution evaluation than Moschopoulos' method.

The exact implementations make it possible to assess the accuracy of the approximation approach of Barnabani (2017). The differences between the approximation and exact evaluation in the two worst cases out of a collection of settings we experimented are shown in Figure 8. Apparently, the two exact methods give indistinguishable results and the approximation is quite accurate.

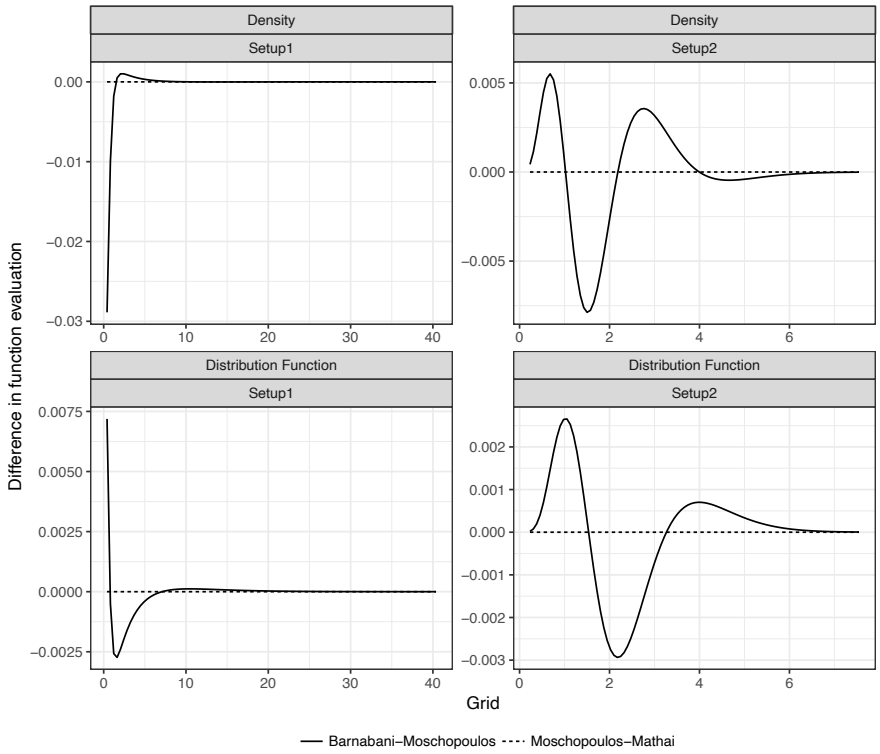

Figure 8: Differences between the approximation method and the exact methods in evaluating the density and distribution function of convolution of three independent gamma variables. The parameters in setup 1 are $\alpha_1 = \alpha_2 = \alpha_3 = 0.2$, $\beta_1 = 4$, $\beta_2 = 0.3$, and $\beta_3 = 0.2$. The parameters in setup 2 are $\alpha_1 = \alpha_2 = \alpha_3 = 2$, $\beta_1 = 0.4$, $\beta_2 = 0.3$, and $\beta_3 = 0.2$.

A.3 Application to a Renewal Process

In many applications, the gamma variables in a convolution of interest are Erlang variables coming from sums of independent and identically distributed exponential variables. Let $\{U_k\}_{k\geq 1}$ be independent and identically distributed random variables following a mixture of two exponential distributions, $\mathcal{E}xp(\beta_1)$ and $\mathcal{E}xp(\beta_2)$, with weights $p \in (0,1)$ and $1-p$, respectively. For any $t > 0$, define

$$N(t) = \sup\{n \geq 0 : \sum_{k=1}^{n} U_k \leq t\}, \tag{A.7}$$

where, by convention, the summation over an empty set is 0. The process $N(t)$, $t > 0$, is a renewal process, and $N(t)$ represents the number of events (renewals) by time t. This process is a special case of the alternating renewal process studied in Crimaldi et al. (2013), where $N(t)$ is used to model a telegrapher's process that is a subject of an additional uncertainty: the switching to an alternating state is random, not deterministic. Our motivating example, however, is from a moving-resting process for animal movement modeling with two different types of resting states (Pozdnyakov et al., 2020).

The following result gives an expression for the distribution of $N(t)$ in terms of the Kummer's confluent hypergeometric function.

Proposition 1. *For the renewal process defined in (A.7) and an integer $n \geq 0$,*

$$\Pr\big(N(t)=n\big) = \sum_{k=0}^{n} \big\{ p\psi(\beta_1,\beta_2,t,k,n) + (1-p)\psi(\beta_2,\beta_1,t,n-k,n) \big\} \times \frac{t^n}{\beta_1^k \beta_2^{n-k}\Gamma(n+1)} \binom{n}{k} p^k (1-p)^{n-k},$$

where

$$\psi(\beta_1,\beta_2,t,k,n) = e^{-t/\beta_1}{}_1F_1\big(n-k; n+1; t(1/\beta_1 - 1/\beta_2)\big).$$

Proof. First, note that

$$\Pr(N(t)=n) = \Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n+1} U_k > t\right)$$
$$= p\Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n} U_k + E_1 > t\right) + (1-p)\Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n} U_k + E_2 > t\right),$$

where E_1 and E_2 are independent of $\{U_k\}_{k\geq 1}$ with $\mathcal{E}xp(\beta_1)$ and $\mathcal{E}xp(\beta_2)$ distributions, respectively. Next, using mixture randomization we get

$$\Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n} U_k + E_1 > t\right) = \Pr\left(\sum_{k=1}^{n} U_k \leq t\right) - \Pr\left(\sum_{k=1}^{n} U_k + E_1 \leq t\right)$$
$$= \sum_{k=0}^{n} H(t;(k,\beta_1),(n-k,\beta_2)) \binom{n}{k} p^k (1-p)^{n-k}.$$

Similarly,

$$\Pr\left(\sum_{k=1}^{n} U_k \le t, \sum_{k=1}^{n} U_k + E_2 > t\right) = \Pr\left(\sum_{k=1}^{n} U_k \le t\right) - \Pr\left(\sum_{k=1}^{n} U_k + E_2 \le t\right)$$
$$= \sum_{k=0}^{n} H(t; (n-k, \beta_2), (k, \beta_1)) \binom{n}{k} p^k (1-p)^{n-k}.$$

That is

$$\Pr(N(t) = n)$$
$$= \sum_{k=0}^{n} \left[pH(t; (k, \beta_1), (n-k, \beta_2)) + (1-p)H(t; (n-k, \beta_2), (k, \beta_1))\right] \binom{n}{k} p^k (1-p)^{n-k} \tag{A.8}$$

Equation (A.9) of Lemma 1 completes the proof. □

Lemma 1. *For any positive integers α_1, α_2 and $y, \beta_1, \beta_2 > 0$,*

$$H(y; (\alpha_1, \beta_1), (\alpha_2, \beta_2)) = \frac{y^{\alpha_1+\alpha_2} e^{-y/\beta_1}}{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \Gamma(\alpha_1 + \alpha_2 + 1)} {}_1F_1(\alpha_2; \alpha_1 + \alpha_2 + 1; y(1/\beta_1 - 1/\beta_2)), \tag{A.9}$$

where $H(x; \alpha_1, \beta_1, \alpha_2, \beta_2) = F(x; \alpha_1, \beta_1, \alpha_2, \beta_2) - F(x; \alpha_1 + 1, \beta_1, \alpha_2, \beta_2)$.

Proof. Equation (A.6) gives us that

$$H(y; (\alpha_1, \beta_1), (\alpha_2, \beta_2))$$
$$= \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} \sum_{k=0}^{\infty} \binom{\alpha_2 + k - 1}{k} (1 - \beta_1/\beta_2)^k G(y/\beta_1; k + \alpha_1 + \alpha_2)$$
$$- \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} \sum_{k=0}^{\infty} \binom{\alpha_2 + k - 1}{k} (1 - \beta_1/\beta_2)^k G(y/\beta_1; k + \alpha_1 + \alpha_2 + 1)$$

$$= \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} \sum_{k=0}^{\infty} \binom{\alpha_2 + k - 1}{k} (1 - \beta_1/\beta_2)^k \frac{(y/\beta_1)^{k+\alpha_1+\alpha_2} e^{-y/\beta_1}}{\Gamma(k + \alpha_1 + \alpha_2 + 1)}$$
$$= \left(\frac{\beta_1}{\beta_2}\right)^{\alpha_2} \left(\frac{y}{\beta_1}\right)^{\alpha_1+\alpha_2} e^{-y/\beta_1} \sum_{k=0}^{\infty} \frac{\binom{\alpha_2+k-1}{k} [y(1/\beta_1 - 1/\beta_2)]^k}{\Gamma(k + \alpha_1 + \alpha_2 + 1)},$$

where the second equation follows from

$$G(y; \alpha) - G(y; \alpha + 1) = \frac{y^\alpha e^{-y}}{\Gamma(\alpha + 1)}.$$

Because of the identities

$$\binom{\alpha_2 + k + 1}{k} = \frac{(\alpha_2)_k}{k!}$$

and

$$\Gamma(k + \alpha_1 + \alpha_2 + 1) = (\alpha_1 + \alpha_2 + 1)_k \Gamma(\alpha_1 + \alpha_2 + 1),$$

we obtain that

$$\begin{aligned} &H(y; (\alpha_1, \beta_1), (\alpha_2, \beta_2)) \\ &\quad = \frac{y^{\alpha_1+\alpha_2} e^{-y/\beta_1}}{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \Gamma(\alpha_1 + \alpha_2 + 1)} \sum_{k=0}^{\infty} \frac{(\alpha_2)_k [y(1/\beta_1 - 1/\beta_2)]^k}{(\alpha_1 + \alpha_2 + 1)_k k!} \\ &\quad = \frac{y^{\alpha_1+\alpha_2} e^{-y/\beta_1}}{\beta_1^{\alpha_1} \beta_2^{\alpha_2} \Gamma(\alpha_1 + \alpha_2 + 1)} {}_1F_1(\alpha_2; \alpha_1 + \alpha_2 + 1; y(1/\beta_1 - 1/\beta_2)). \end{aligned}$$

□

This result has its own value. The difference of distribution functions of convolution of two independent Erlang distributions with shape parameters that differ by 1

plays an important role in computation of distribution of the occupation times for a certain continuous time Markov chain (Pozdnyakov et al., 2020), the defective density in generalized integrated telegraph processes (Zacks, 2004), and the first-exit time in a compound process (Perry et al., 1999).

Finally, the proof of Proposition 1 can be extended to a mixture of more than two exponential random variables. But the resulting formula is computationally heavy because it does not benefit from the usage of special functions.

Corollary 1. *Let* $\{U_k\}_{k\geq 1}$ *be independent and identically distributed random variables distributed as a mixture of* S *exponential distributions,* $\mathcal{E}xp(\beta_s)$ *with weight* p_s, $s = 1,\ldots,S$, *and* $\sum_{s=1}^{S} p_s = 1$. *For* $n \geq 0$,

$$\begin{aligned}&\Pr\left(N(t)=n\right)\\ &=\sum_{k_1=0}^{n}\sum_{k_2=0}^{n-k_1}\cdots\sum_{k_{S-1}=0}^{n-k_1-\cdots-k_{S-2}}\left[\left(\sum_{s=1}^{S}p_s H_s(t;(k_1,\beta_1),\ldots,(k_S,\beta_S))\right)\frac{p_1^{k_1}\ldots p_S^{k_S}}{k_1!k_2!\ldots k_S!}n!\right],\end{aligned}$$

where

$$\begin{aligned}H_s(t;(k_1,\beta_1),\ldots,(k_S,\beta_S)) &= F_S(t;(k_1,\beta_1),\ldots,(k_s,\beta_s),\ldots,(k_S,\beta_S))\\ &\quad - F_S(t;(k_1,\beta_1),\ldots,(k_s+1,\beta_s),\ldots,(k_S,\beta_S)),\end{aligned}$$

$s = 1,\ldots,S$, $k_S = n - k_1 - \cdots - k_{S-1}$, *and* F_S *is the distribution function of the convolution of* S *independent gamma variables with shape parameter* $(k_1,\ldots,k_S)$ *and*

scale parameter $(\beta_1, \ldots, \beta_S)$.

Proof. Note first that

$$\begin{aligned}\Pr(N(t) = n) &= \Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n+1} U_k > t\right) \\ &= \sum_{s=1}^{S} p_s \Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n} U_k + E_s > t\right),\end{aligned}$$

where E_s are independent of $\{U_k\}_{k\geq 1}$ with an $\mathcal{E}xp(\beta_s)$ distribution. Then we get that

$$\begin{aligned}&\Pr\left(\sum_{k=1}^{n} U_k \leq t, \sum_{k=1}^{n} U_k + E_s > t\right) \\ &= \Pr\left(\sum_{k=1}^{n} U_k \leq t\right) - \Pr\left(\sum_{k=1}^{n} U_k + E_s \leq t\right) \\ &= \sum_{k_1=0}^{n} \sum_{k_2=0}^{n-k_1} \cdots \sum_{k_{S-1}=0}^{n-k_1-\cdots-k_{S-2}} H_s(t, (k_1, \beta_1), \ldots, (k_S, \beta_S)) \frac{n!}{k_1! k_2! \ldots k_S!} p_1^{k_1} \ldots p_S^{k_S}.\end{aligned}$$

□

To demonstrate the computational efficiency of the result of Proposition 1, we performed a numerical study for $S = 2$. The scale parameters of the exponential distributions were set to be $(\beta_1, \beta_2) \in \{(0.4, 0.3), (4, 0.3), (4, 3)\}$. For each configuration, we evaluated $\Pr(N(t) = n)$ for $t = 10$ and three n values corresponding to the 20th, 50th, and 90th percentiles of a random sample of size 1,000 from the distribution of $N(10)$. The H function was evaluated with Mathai's method, Moschopoulos' method and the proposed result in Proposition 1. The numerical results are identical and the median

Table 10: Timing comparison (in microseconds) using different methods to evaluate $\Pr(N(10) = n)$ in the renewal process application when $S = 2$.

Parameters			Time (microseconds)		
β_1	β_2	n	Mathai	Moschopoulos	Proposed
0.4	0.3	27	4,993	36,314	33
		32	5,788	43,129	34
		40	6,981	54,637	41
4	0.3	10	2,675	23,613	15
		18	4,667	42,529	23
		30	7,525	65,408	58
4	3	2	308	951	9
		3	380	1,497	15
		5	556	3,226	16

timing results from 100 replicates are summarized in Table 10. The method in Proposition 1, which bypasses evaluating two distribution function of the convolution of two exponential variables, shows a significant advantage, speeding up Mathai's method by a factor in the range of 60–200.

For $S > 2$, we performed a similar study with Mathai's and Moschopoulos' method for evaluating H_3 using the result in Corollary 1. For comparison in accuracy and speed, the approximation method from Barnabani (2017) in evaluating H_3 was also included. The scale parameters of the exponential distributions were set to be

$$(\beta_1, \beta_2, \beta_3) \in \{(0.4, 0.3, 0.2),\ (4, 0.3, 0.2),\ (4, 3, 0.2),\ (4, 3, 2)\}.$$

Table 11: Timing comparison (in milliseconds) using different methods to evaluate $\Pr(N(10) = n)$ in the renewal process application when $S = 3$.

Parameters				Time (milliseconds)			Exact Value	Relative Error
β_1	β_2	β_3	n	Mathai	Moschopoulos	Barnabani	($\times 10^{-2}$)	(‰)
0.4	0.3	0.2	36	95,440	3,033	669	4.26	3.39
			42	132,284	4,097	895	5.76	−2.58
			51	189,053	5,901	1,285	2.28	−0.48
4	0.3	0.2	10	8,729	376	63	2.83	3.67
			19	33,312	1,249	218	3.40	−0.03
			35	113,068	3,906	674	1.49	−10.63
4	3	0.2	5	2,791	94	15	5.89	0.33
			10	12,937	382	62	6.28	−2.07
			19	49,028	1,229	203	2.12	1.52
4	3	2	2	31	6	1	12.85	1.22
			4	233	26	3	18.74	−2.00
			7	829	62	11	7.21	1.98

Again, for each configuration, we evaluated $\Pr(N(10) = n)$ for three n values corresponding to the 20th, 50th, and 90th percentiles of a random sample of size 1000 from the distribution of $N(10)$. The median timing results from 100 replicates and the relative error of the approximation method are summarized in Table 11. Similar to the comparison reported in Section A.2, Moschopoulos' method is over 10 times faster than Mathai's method in all the cases. A small summation of β's requires longer computation time. The approximation method is over 6 times faster than Moschopoulos' method, with relative error less than 1 percent.

A.4 Discussion

We reviewed and compared two exact methods and one approximation method for evaluating the density and distribution of convolutions of independent gamma variables. From our study, the method of Mathai (1982) is the fastest in the case of $n = 2$ because of the efficient GSL implementation of the univariate Kummer's confluent hypergeometric function; when $n > 3$, the method of Moschopoulos (1985) is faster than Mathai's method. The fast approximation method of Barnabani (2017) is quite accurate, which provides a useful tool for applications with $n > 3$.

One important application arises when gamma variables are Erlang variables resulting from sums of independent and identically distributed exponential variables. Proposition 1 provides the exact distribution of the event count in the renewal process with holding time following a mixture of exponential distributions. This result allows a fast numerical evaluation because of the usage of Kummer's confluent hypergeometric functions.

Implementations of the exact methods of Mathai (1982) and Moschopoulos (1985), as well as the approximation method of Barnabani (2017), are available in our open-source R package `coga` (Hu et al., 2019), which is built on C++ code for fast speed. The method of Di Salvo (2008) is also implemented, which is based on multiple integrals and, hence, is slower than that of Moschopoulos (1985). The package implementation fills a gap in the computing tools available to practitioners in a variety of applications

involving convolutions of independent gamma variables. Further improvement will be possible if efficient implementation of special functions such as the multivariate confluent hypergeometric function and the fourth Lauricella function.

A related topic is matrix variate gamma distribution (Gupta and Nagar, 2018). Considerable interest of matrix variate gamma arises in Bayesian statistics. An example is the Wishart distribution, a special case of matrix variate gamma. Known as the distribution of the sample covariance matrix of multivariate normal distributions, it is the conjugate prior of the precision matrix in the Gaussian case. The properties of matrix variate gamma distribution and the inverted matrix variate gamma have been studied (Mathai, 2005; Iranmanesh et al., 2013). Consequently, it may also be of interest to work out the distribution of convolution of matrix variate gamma variables, this can be defined according to the convolution of functions on a positive definite matrix space. The corresponding exact density was given by Di Salvo (2008) in term of matrix integral, but the numerical evaluation is extremely challenging.

Bibliography

Abramowitz, M. and I. A. Stegun (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Volume 9. Dover, New York.

Akkouchi, M. (2005). On the convolution of gamma distributions. *Soochow Journal of Mathematics 31*(2), 205–211.

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72*(3), 269–342.

Barnabani, M. (2017). An approximation to the convolution of gamma distributions. *Communications in Statistics - Simulation and Computation 46*(1), 331–343.

Benhamou, S. (2011, 01). Dynamic approach to space and habitat use based on biased random bridges. *PLoS ONE 6*(1), e14592.

Bon, J.-L. and E. Pãltãnea (1999). Ordering properties of convolutions of exponential random variables. *Lifetime Data Analysis 5*(2), 185–192.

Bshouty, D., A. Di Crescenzo, B. Martinucci, and S. Zacks (2012). Generalized telegraph process with random delays. *Journal of Applied Probability 49*, 850–865.

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in Hidden Markov Models*. Springer.

Codling, E. A., M. J. Plank, and S. Benhamou (2008). Random walk models in biology. *Journal of The Royal Society Interface 5*(25), 813–834.

Crimaldi, I., A. Di Crescenzo, A. Iuliano, and B. Martinucci (2013). A generalized telegraph process with velocity driven by random trials. *Advances in Applied Probability 45*(4), 1111–1136.

De Gregorio, A. and S. M. Iacus (2008). Parametric estimation for the standard and geometric telegraph process observed at discrete times. *Statistical Inference for Stochastic Processes 11*(3), 249–263.

De Gregorio, A. and S. M. Iacus (2011). Least-squares change-point estimation for the telegraph process observed at discrete times. *Statistics 45*(4), 349–359.

Di Crescenzo, A. (2001). On random motions with velocities alternating at Erlang-distributed random times. *Advances in Applied Probability 33*, 690–701.

Di Crescenzo, A., B. Martinucci, and S. Zacks (2014). On the geometric brownian motion with alternating trend. *Mathematical and statistical methods for actuarial sciences and finance, Eds. Perna, C. and Sibillo, M.*, 81–85.

Di Crescenzo, A. and F. Pellerey (2002). On prices' evolutions based on geometric telegrapher's process. *Applied Stochastic Models in Business and Industry. 18*(2), 171–184.

Di Crescenzo, A. and S. Zacks (2015). Probability law and flow function of Brownian motion driven by a generalized telegraph process. *Methodology and Computing in Applied Probability 17*(3), 761–780.

Di Salvo, F. (2008). A characterization of the distribution of a weighted sum of gamma variables through multiple hypergeometric functions. *Integral Transforms and Special Functions 19*(8), 563–575.

Efron, B. and D. Hinkley (1978). Assessing the accuracy of the maximum likelihood estimator: observed versus expected Fisher information. *Biometrika 65*, 457–487.

Efron, B. and R. Tibshirani (1994). *An Introduction to the Bootstrap*. CRC Press.

Furman, E. and Z. Landsman (2005). Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics 37*(3), 635–649.

Galassi, M., J. Davies, J. Theiler, B. Gough, and G. Jungman (2009). *GNU Scientific Library: Reference Manual* (3 ed.). Network Theory Ltd.

Godambe, V. P. (1960, 12). An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist. 31*(4), 1208–1211.

Gupta, A. K. and D. K. Nagar (2018). *Matrix Variate Distributions*. Chapman and Hall/CRC.

Horne, J. S., E. O. Garton, S. M. Krone, and S. Lewis, J (2007). Analyzing animal movements using Brownian bridges. *Ecology 88*(9), 2354–2363.

Hu, C., V. Pozdnyakov, and J. Yan (2019). *coga: Convolution of Gamma Distributions*. R package version 1.0.0.

Hu, C., V. Pozdnyakov, and J. Yan (2020a). Density and distribution evaluation for convolution of independent gamma variables. *Computational Statistics 35*(1), 327–342.

Hu, C., V. Pozdnyakov, and J. Yan (2020b). *smam: Statistical Modeling of Animal Movement*. R package version 0.5.3.

Iacus, S. M. and N. Yoshida (2008). Estimation for the discretely observed telegraph process. *Teorīya Ĭmovīrnosteĭ ta Matematichna Statistika 78*, 32–42.

Ionides, E. L., A. Bhadra, Y. Atchadé, and A. King (2011). Iterated filtering. *The Annals of Statistics 39*(3), 1776–1802.

Ionides, E. L., D. Nguyen, Y. Atchadé, S. Stoev, and A. A. King (2015). Inference for dynamic and latent variable models via iterated, perturbed Bayes maps. *Proceedings of the National Academy of Sciences 112*(3), 719–724.

Iranmanesh, A., M. Arashi, D. Nagar, and S. Tabatabaey (2013). On inverted matrix variate gamma distribution. *Communications in Statistics-Theory and Methods 42*(1), 28–41.

Jain, G. C. and P. C. Consul (1971). A generalized negative binomial distribution. *SIAM Journal on Applied Mathematics 21*(4), 501–513.

Jasiulewicz, H. and W. Kordecki (2003). Convolutions of Erlang and of Pascal distributions with applications to reliability. *Demonstratio Mathematica. Warsaw Technical University Institute of Mathematics 36*(1), 231–238.

Kaas, R., M. Goovaerts, J. Dhaene, and M. Denuit (2008). *Modern Actuarial Risk Theory: Using R*, Volume 128. Springer Science & Business Media.

Kadri, T., K. Smaili, and S. Kadry (2015). Markov modeling for reliability analysis using hypoexponential distribution. In S. Kadry and A. El Hami (Eds.), *Numerical Methods for Reliability and Safety Assessment: Multiscale and Multiphysics Systems*, pp. 599–620. Cham: Springer.

Khaledi, B.-E. and S. Kochar (2013). A review on convolutions of gamma random variables. In H. Li and X. Li (Eds.), *Stochastic Orders in Reliability and Risk*, pp. 199–217. Springer.

King, A. A., D. Nguyen, and E. L. Ionides (2016). Statistical inference for partially observed Markov processes via the R package pomp. *Journal of Statistical Software 69*(12), 1–43.

Kolesnik, A. D. and N. Ratanov (2013). *Telegraph processes and option pricing*. Springer Briefs in Statistics. Springer, Heidelberg.

Kranstauber, B., R. Kays, S. D. LaPoint, M. Wikelski, and K. Safi (2012). A dynamic brownian bridge movement model to estimate utilization distributions for heterogeneous animal movement. *Journal of Animal Ecology 81*(4), 738–746.

Lindsay, B. G. (1988). Composite likelihood methods. *Contemporary Mathematics 80*, 221–239.

Mathai, A. (2005). A pathway to matrix-variate gamma and normal densities. *Linear Algebra and Its Applications 396*, 317–328.

Mathai, A. M. (1982). Storage capacity of a dam with gamma type inputs. *Annals of the Institute of Statistical Mathematics 34*(1), 591–597.

Mathai, A. M. and R. K. Saxena (1978). *The H function with Applications in Statistics and Other Disciplines*. New Delhi: Wiley.

Moschopoulos, P. G. (1985). The distribution of the sum of independent gamma random variables. *Annals of the Institute of Statistical Mathematics 37*(1), 541–544.

Nadarajah, S. and S. Kotz (2007). On the convolution of pareto and gamma distributions. *Computer Networks 51*(12), 3650–3654.

Norris, J. R. (1998). *Markov Chains*, Volume 2 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. Reprint of 1997 original.

Othmer, H. G., S. R. Dunbar, and W. Alt (1988). Models of dispersal in biological systems. *Journal of Mathematical Biology 26*(3), 263–298.

Owari, T., H. Kasahara, N. Oikawa, and S. Fukuoka (2009, 01). Seasonal variation of global positioning system (gps) accuracy within the tokyo university forest in hokkaido. *Bulletin of Tokyo University Forests 120*, 19–28.

Perry, D., W. Stadje, and S. Zacks (1999). First-exit times for increasing compound processes. *Communications in Statistics: Stochastic Models 15*(5), 977–992.

Pozdnyakov, V., L. Elbroch, C. Hu, T. Meyer, and J. Yan (2020). On estimation for brownian motion governed by telegraph process with multiple off states. *Methodology and Computing in Applied Probability*.

Pozdnyakov, V., L. Elbroch, A. Labarga, T. Meyer, and J. Yan (2019, Sep). Discretely observed brownian motion governed by telegraph process: Estimation. *Methodology and Computing in Applied Probability 21*(3), 907–920.

Pozdnyakov, V., T. H. Meyer, Y.-B. Wang, and J. Yan (2014). On modeling animal movements using brownian motion with measurement error. *Ecology 95*, 247–253.

Preisler, H. K., A. A. Ager, B. K. Johnson, and J. G. Kie (2004). Modeling animal movements using stochastic differential equations. *Environmetrics 15*(7), 643–657.

R Core Team (2018). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sen, A. and N. Balakrishnan (1999). Convolution of geometrics and a reliability problem. *Statistics & Probability Letters 43*(4), 421–426.

Sericola, B. (2000). Occupation times in markov processes. *Communications in Statistics. Stochastic Models 16*(5), 479–510.

Sim, C. (1992). Point processes with correlated gamma interarrival times. *Statistics & Probability Letters 15*(2), 135–141.

Srivastava, H. M. and P. W. Karlsson (1985). *Multiple Gaussian Hypergeometric Series*. Ellis Horwood Series in Mathematics and Its Applications. E. Horwood.

Stadje, W. and S. Zacks (2004). Telegraph processes with random velocities. *Journal of Applied Probability 41*, 665–678.

Tilles, P. F. C. and S. V. Petrovskii (2016). How animals move along? exactly solvable model of superdiffusive spread resulting from animal's decision making. *Journal of Mathematical Biology 73*, 227–55.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*(1), 5–42.

Vellaisamy, P. and N. Upadhye (2009). On the sums of compound negative binomial and gamma random variables. *Journal of Applied Probability 46*(1), 272–283.

Viterbi, A. J. (2006). A personal history of the Viterbi algorithm. *IEEE Signal Processing Magazine 23*(4), 120–142.

Yan, J., Y.-W. Chen, K. Lawrence-Apfel, I. Ortega, V. Pozdnyakov, S. Williams, and T. Meyer (2014). A moving-resting process with an embedded Brownian motion for animal movements. *Population Ecology 56*(2), 401–415.

Zacks, S. (2004). Generalized integrated telegraph processes and the distribution of related stopping times. *Journal of Applied Probability 41*(2), 497–507.

Zacks, S. (2012). Distribution of the total time in a mode of an alternating renewal process with applications. *Sequential Analysis 31*(3), 397–408.

Zucchini, W., MacDonald, I. L., and R. Langrock (2016). *Hidden Markov Models for Time Series: An Introduction Using R* (2 ed.). Chapman and Hall/CRC.